NOVELTY, PREDICTION ERROR AND MEMORY ENCODING: LIMITATIONS OF THE PIMMS FRAMEWORK

Jörn Alexander Quent

Clare Hall

MRC Cognition and Brain Sciences Unit

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

November 2021

DECLARATION

This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

Chapter 1 is in parts based on a review article by Quent, J. A., Henson, R. N., & Greve, A. (2021). A predictive account of how novelty influences declarative memory. *Neurobiology of Learning and Memory*, 179, 107382.

Parts of Chapter 1 and Chapter 3 are based on a registered report by Quent, J. A. and Henson, R. N. (accepted for Stage 1) *Quarterly Journal of Experimental Psychology*.

This thesis does not exceed the prescribed word limit for the relevant Degree Committee.

Jörn Alexander Quent, B.Sc. & M.Sc.

Beijing, People's Republic of China.

ABSTRACT

The Predictive Interactive Multiple Memory Systems (PIMMS) framework has been used to explain how novelty, or more precisely "prediction error", boosts memory encoding. In this thesis, I explored several other phenomena in the animal and human literature that PIMMS cannot yet explain but should.

PIMMS predicts that unexpected information will be better encoded than expected information. However recent work has suggested that expected information can also be better remembered than less expected information. By using a range of expectancies for the location of objects with an immersive virtual reality (iVR) kitchen, I showed that memory is a "U-shaped" function of expectancy, with best memory for highly expected or highly unexpected locations relative to intermediate levels of expectancy. Using OSF-registered Bayesian inference, this U-shape was consistent across four experiments. While the advantage for highly unexpected locations is consistent with PIMMS, the advantage for highly expected locations is not. Importantly, the advantage for expected locations was not simply due to a guessing bias when the location was forgotten, suggesting that the advantage arises during encoding rather than just at retrieval.

This U-shape is consistent with another framework - the SLIMM framework - which proposes that different brain regions support the two ends of the U-shape, such that the advantage for unexpected information should be associated with recollection of contextual information via a medial temporal lobe system (like in PIMMS), while the advantage for expected information should be associated with a feeling of familiarity based on rapid cortical consolidation enabled by a medial prefrontal cortex system. However, when I asked participants to indicate recollection or familiarity at retrieval, both ends of the U-shape continuum were associated with higher recollection, while there was no detectable effect of expectancy on familiarity. I consider why this SLIMM prediction may therefore be incorrect.

Another finding in the literature concerns the effect of novelty on unrelated information shortly preceding or succeeding the novel experience. PIMMS says nothing about this penumbra effect, which has been related to plasticity-related proteins triggered by the novel experience (so-called "behavioural tagging"). Since participants report that their first iVR experience is highly novel, I submitted a Registered Report to test whether iVR affected memory for unrelated words that were encountered prior to entering the iVR

room. In short, the finding was that there is no evidence that novelty improves memory performance for information learned before experiencing something novel. Possible reasons for the failure of finding an effect were discussed.

A final limitation of PIMMS I considered was the effect of “boundaries” in continuous stimuli, which are known to affect memory for the temporal order of information. While boundaries might be generated by prediction errors, PIMMS is silent on how they affect temporal order memory. Using a movie featuring a series of rooms, I tested whether memory for the temporal order of objects encountered in those rooms is affected by doorways between rooms and/or by surprising/perceptual changes within a room. Unfortunately, I was unable to replicate a previous report where temporal order memory was worse for pairs of objects in different rooms (i.e., either side of a doorway) than objects in the same room, let alone either sides of a surprising/perceptual change within a room.

Taken together, my findings and the literature demonstrate the multiple potential factors that determine how novelty affects memory encoding (and consolidation), which require a more comprehensive theoretical framework than currently available.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Professor Rik Henson, who guided me on this journey while giving me freedom to work on the things that interest me, offering advice and sometimes where necessary reigning in my haste. If I had to pick one thing that I valued the most, then it is his fair but very honest criticism, which helped me grow. I hope that completing my PhD will not be the last time that I can work with such a great scientist and mentor.

This PhD thesis would have been impossible without the generous support that was given to me by the Gates Cambridge Trust as without the help of this wonderful lab that I was fortunate enough to be part of. I want to therefore thank past and present members directly: Dr. Andrea Greve, Dr. Aya Ben-Yakov, Dr. Roni Tibon, Dr. Reene Visser, Tina Emery, Dr. Alex Kaula, Levan Bokeria. Kevin Campion, Dr. Sung-Mu (Scott) Lee , Dr. Ethan Knights and last but certainly not least Dr. Elisa Cooper. I hope that I will meet each and every one again. I am eternally grateful for this lab; after all they had to put up with me but at the same time they were so patient and welcoming.

The same is true for the incredible kind human beings that I shared the office with. I already miss them and our ‘short’ chats during ‘lunch’ time. My time in Cambridge was immeasurably better through their presence. I could have not asked for better office mates.

I also like to extend my thanks to the CBU as a whole. There are too many people to name individually but without the support of the countless individuals my thesis would have been impossible so they cannot go unacknowledged. I know I had my fair share of complaints about the bureaucracy, rules and regulations that had to be followed but at the end of day (or my PhD) I am grateful for their often unnoticed and tiring efforts that made my life easier. I would like to mention the IT team of the CBU (Howard Gyton, Jeff Berry, Anthony Parry-Jones and Dr. Russell Thompson) at this point whose support I greatly valued.

Special thanks goes to my family and especially my parents who tried as best as they could to give me the support and mental nourishment and yet had to see me drop out of school without a degree. I can only imagine the pain and fear for my future that this has caused them. I hope now they can enjoy that their child finished his PhD at the University of Cambridge against all odds.

For most of my high-school career, I was an underperforming student. Yet, there were always teachers who recognised my potential and stood by me. Therefore, I want to especially thank the teachers at the Fichte Gymnasium Hagen and Rahel Varnhagen Kolleg Hagen without whom my path most certainly would have taken a completely different direction. Mrs. Sonnenschein and many others deserve credit for not giving up on me.

Finally, I would like to finish by thanking Dr. Mengya Zhang. I take great pride in the honour of calling you my wife. More than this PhD, you are the greatest gift the CBU and Cambridge have given me. I cannot wait to spend the future with you.

CONTENTS

LIST OF TABLES

List of Figures

LIST OF APPENDICES

1 INTRODUCTION

Events that are new, different or unusual often “stick in our minds”. For example, imagine you live and work in an urban area and you take the same route to work in your car every morning. One day you see a flock of sheep blocking the road on your commute. You will probably remember this event for a long time, while the other countless times you have driven down that road will be inaccessible. Numerous studies have confirmed this observation, namely that if we experience a novel event in a familiar context, we tend to store and remember this event more easily. However, if you happen to live in the countryside close to a sheep farm, your experience might be quite different: because you see sheep quite frequently, you might experience just another regular commute to work that is hardly memorable. Thus, the experience of novelty is not ‘absolute’ and cannot be defined independent of the observer; rather it is driven by what an individual expects to experience compared to what they actually encounter. This comparison between expectation and experience parallels the computation of a prediction error (PE) in many other theories of learning. According to such theories, we continuously generate expectations about our environment and update those predictions when they are wrong, i.e., when a PE occurs. While the role of PE in learning is well established in experiments on certain types of memory (e.g., in conditioning and associative learning; Rescorla & Wagner, 1972), its role in conscious and verbalisable memory (i.e. declarative memory; see below for explanation) is less so. The theoretical framework that is central to this thesis, which helps to shed some light on the complex relationship between novelty, PE and memory (encoding), is the ‘Predictive Interactive Multiple Memory Signals’

(PIMMS) proposal of Henson & Gagnepain (2010). While this powerful framework helps us to understand – among other things – how novelty relates to declarative memory, it also has several limitations. The topic of my thesis thus is to highlight and examine empirically some of those shortcomings. First, I will briefly recap major theories of memory, in order to relate them to the scope of this thesis.

1.1 (Very) Brief Review of Human Memory

Human memory is widely believed to not be a unitary construct, but rather contains numerous distinctions. One of the most notable distinctions is between short-term memory (STM), which stores information temporarily for a matter of seconds, and long-term memory (LTM), which is generally believed to be able to store some information indefinitely. This distinction goes back to at least Atkinson and Shiffrin (1968) and is still upheld today, even though some disagree (see Norris, 2017, for a review on this question). In my thesis, I focus on LTM, but there are potential interactions with STM (or "working memory", WM) in the final empirical chapter on event segmentation.

Another important distinction is between declarative memory (i.e., conscious, verbalisable memory) and non-declarative memory (e.g., procedural memory; Squire, 2004). Central to this thesis, the former type has been subdivided into semantic memory ("general knowledge of the world") and episodic memory ("remembering personally experienced past"; Tulving, 1995, p. 841), while others have proposed an additional "perceptual memory" system that underlies non-declarative phenomena such as perceptual priming (Schacter & Tulving, 1994). In his seminal SPI model, Tulving (1995) explains encoding, storage and retrieval in perceptual, semantic and episodic memory systems as *serial*, *parallel* and *independent* (hence the name SPI): information is encoded serially, by passing from perceptual through semantic to episodic memory systems, which leaves multiple potential memory traces in each system in parallel, any of which can be retrieved independently (depending on the task). The PIMMS framework questioned this model however, accepting that experiences can leave memory traces in multiple levels throughout the brain, but claiming that interactions between these levels are important for both encoding and retrieval (an "interactive-parallel-interactive" model of encoding-storage-retrieval). These interactions take the specific form of (top-down) predictions and (bottom-up) PE across a hierarchical organisation of these levels (see later). Moreover, those levels are defined by the type of representations they process (rather than a specific memory function), which might be perceptual features (e.g. colours and shapes), items (e.g., everyday objects defined by their perceptual features) and context (e.g., spatiotemporal environment in which objects are encountered). While there are no doubt many more such levels in reality, these three allow definition of episodic memory (that can lead to *recollection*) - as new associations between contexts and items - and semantic memory (that can lead to *familiarity*) - as (temporary changes) in existing associations between items and their features (Figure 1.1). PIMMS tentatively maps these three levels onto different brain regions, though these are only indirectly relevant to the present thesis.

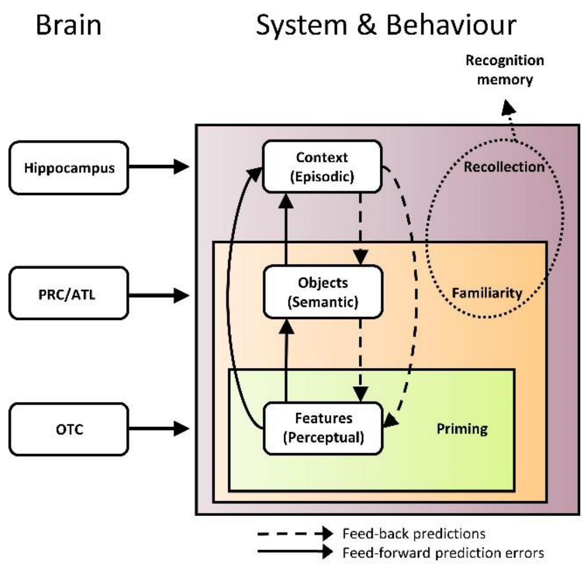

Figure 1.1: Schematic illustration of the PIMMS framework: Illustrated here is the PIMMS model that shows how different types of content (features, objects and contexts) are represented layers that roughly map onto SPI's memory systems (perceptual, semantic and episodic). Dashed and solid lines represent feedback predictions and feedforward PE. Retrieval from the perceptual, semantic and episodic memory manifest in priming, familiarity and recollection respectively. The latter two are both contributing to recognition memory. Furthermore, on the left shown are brain regions that are postulated to mainly contributed to the respective layers. PRC = perirhinal cortex, ATL = anterior temporal lobe and OTC = occipital-temporal cortex.

Chapter 2 deals with PIMMS's (in)ability to explain how well new associations between items and their (spatial) context are encoded as a function of the degree of PE.

For Chapter 3, it is necessary to further introduce the process of *consolidation*, in addition to encoding and retrieval. Consolidation refers to the time period between encoding and retrieval, during which a memory is not necessarily passively stored in some latent state, but is actively transformed from a temporary to a more permanent state, which may involve specialised brain mechanisms, e.g., during sleep. Indeed, consolidation likely involves multiple processes at different spatial and temporal scales (gene expression, synaptic remodelling, etc.), and an important distinction has been made between *cellular/synaptic consolidation* and *system consolidation*. The former describes physiological processes that change synaptic strengths by inducing structural changes in cells, which in principle might occur throughout the brain. This type of consolidation is less contentious than *system consolidation* (Dudai, 2004; Frankland & Bontempi, 2005), which refers to the time-dependent role of specific brain regions in memory, such as the medial temporal lobe (MTL) in episodic memory. The standard theory of system consolidation posits that new memory traces are initially dependent on the hippocampus until they are gradually transferred to the neocortex; a process that might take years to occur (Frankland & Bontempi, 2005). Note however that other theories, such as “multiple trace theory” (Nadel & Moscovitch, 1997), argue that detailed memories always depend on the hippocampus and each time a memory is encoded, another memory trace is created, explaining why memories can get stronger and more resistant to forgetting. A related theory is the “context binding theory” (Yonelinas et al., 2019), which argues the hippocampus binds item-related and context-related information, that retrieval of detailed item-context information is always hippocampus-dependent and that forgetting is simply the result of contextual interference. In any case, the majority of consolidation theories posit that there is a time window during which memories can be modulated after their initial encoding, such as the occurrence of a very novel/salient event. This is tested in Chapter 3, since PIMMS is largely silent on consolidation, and so needs elaboration.

Finally, most memory experiments use discrete stimuli, whereas real experience is continuous. It has been suggested that memories are formed by segmenting that experience. PE is believed to play a crucial role as has been suggested in various empirical and computational work (e.g. Reynolds et al., 2007). While PIMMS formalises PE, it has not yet been extended to continuous stimuli, but treats an event as a stationary snapshot without any temporal dimension. Moreover, event segmentation affects memory in various ways, such as memory for temporal order, that are not specified by PIMMS. This is tested in the final empirical part (Chapter 4), where I attempted to replicate a study by Horner et al. (2016) that showed that moving through doors in a virtual environment affects memory for the order of objects encountered in those rooms.

1.2 Novelty/surprise and memory encoding in PIMMS

Coming back to PIMMS, higher levels in a perceptual-mnemonic hierarchy are constantly predicting the activity in lower levels and the difference between the predicted and actual activity (at a single moment in time) – the PE – is fed back from lower to higher levels. For a given layer, the predictions from the level above serve as a *prior* probability distribution (in a Bayesian sense), whereas the activity profile (produced from the level below) is equivalent to the evidence or *likelihood,* while the PE is the *divergence* between these two distributions (the summed area of no overlap; see Figure 1.2)[1]. According to PIMMS, the size of the PE then determines how well a new event is encoded into memory. This PE acts "locally", in the sense that it strengthens just the association between just the active (winning) representations in each connected level (similar to the "delta" learning rule in Rescorla & Wagner, 1972), such as between a specific context and an (unexpected) object encountered in that context.

One of the earliest reported examples of novelty/surprise in memory is the *von Restorff effect* (or *isolation effect*). In one of her seminal experiments, von Restorff (1933; for a translation see Hunt, 1995) presented four pairs of syllables along with one pair of symbols, numbers, letters or patches of colour. The atypical pairs were better recalled immediately afterwards. While this seems consistent with PIMMS's hypothesis that PE drives better memory encoding, the isolation effect can also arise for items early in the series, at a point where the typical list composition is not apparent, so no clear predictions should be possible (e.g. see von Restorff, 1933). Therefore, von Restorff herself actually dismissed encoding as the process underlying the memory benefit. Similarly, others argued that von Restorff's effect can be explained with distinctiveness at retrieval (Dunlosky et al., 2000; Kelley & Nairne, 2001). Bruce and Gaines (1976) showed that

[1] More precisely, the PE that drives learning is the divergence between the prior and the posterior, on the assumption that the initial PE between the prior and the likelihood is what drives perception (activity changes), which seeks to minimise this PE over a few hundred milliseconds (producing the posterior distribution), leaving the residual PE that drives synaptic change instead (Henson & Gagnepain, 2010). For simplicity though, I assume here that the posterior is close to the likelihood, such that the qualitative implications are the same.

recall of physically distinct words was clustered, which might suggest that these items are stored together aiding their retrieval. Others posit there is a direct link between distinctiveness and the retrieval context which is to say that it provides an advantage in retrieving items from that specific context (McDaniel et al., 1995).

In contrast, Geraci & Manzano (2010) argued that stimulus salience develops over the course of an episode that is the whole presentation of a list, which allows early list items to benefit from surprise. Furthermore, more recent work (Schmidt & Schmidt, 2017) has also demonstrated that unexpected events benefit from processes that operate not only during retrieval but also during encoding, depending on the context and task demands. So, while some aspects of the *von Restorff effect* might be due to non-PE related retrieval mechanisms, PE, which leads better encoding, undoubtably plays some role in this phenomenon.

A related finding, which has been coined the classic ‘novelty effect’ in episodic memory, was popularised by Tulving and Kroll (1995), though first examined by Kinsbourne and George (1974). In this paradigm, participants are familiarised with a list of random words. In a second “critical” phase, they see a list of new words, intermixed with some of the familiarised words. Finally, they are presented with a third list of words, and asked to recognise any that came from the second, critical phase. A common finding in this paradigm is that the new words are better recognised than the familiar words, despite the fact that they were presented fewer times in total[2]. This led to the formulation of the *novelty-encoding hypothesis*, which postulates that novel information is preferentially encoded by brain regions like the hippocampus (Tulving et al., 1996), which supports episodic memory.

However, the cause of this ‘classic’ novelty effect has been challenged by alternative explanations, other than a special role for novelty in encoding, such as distinctiveness or source confusion at retrieval (Åberg & Nilsson, 2001; Dobbins et al., 1998; Greene, 1999; Poppenk et al., 2010). In general, these authors attribute the difference in recognition to

[2] Note, it is critical that participants are asked to recognise items specifically from the critical phase; if participants are instructed to recognise items studied in either list, memory is better for items that were repeatedly presented (Kafkas & Montaldi, 2015a; Kim et al., 2012).

an impairment for familiar stimuli owing to contextual interference, e.g., “Did I see this stimulus in the familiarisation phase (first list) or study phase (second list)?”, rather than benefit for the non-familiarised (novel) stimuli, which produces increased false alarm rates and hence overall decreased old/new discrimination. However, other studies controlled for confounds like interference, and reinforced Tulving and Kroll’s (1995) conclusion that there is a benefit for this type of words. Åberg and Nilsson (2003), for instance, reported a novelty effect for high confidence responses that showed both increases in hit rates and decreases in false alarm rates for novel items, which suggests the effect cannot be explained fully by reduced retrieval accuracy for familiar items, but rather enhanced encoding of novel items. Similarly, Kormi-Nouri et al. (2005) presented distinct encoding tasks to minimize source confusion at retrieval, and still observed better memory for novel over familiar words, consistent with the general idea of the novelty-encoding hypothesis.

PIMMS’s perspective on this “classic” novelty effect is somewhat different. Its explanation of this effect relates to predictions made by the temporal context: When you encounter the list of random words in the familiarisation phase, you are unable to predict the next word. However, in the critical phase, you notice that some of the words are repeated from the familiarisation phase, so you might start to expect further repetitions. Indeed, if novel items become less expected, they will elicit a greater PE than the familiar words, and therefore become more strongly associated with the context of the critical study phase. Therefore, when finally asked to recognise words specifically from that phase, you are better able to do so. If this interpretation is correct, then the advantage of these “novel” words should depend on the strength of expectation for repeated words, which could be tested by manipulating the ratio of familiarised to new words in the critical phase. This prediction is consistent with Kafkas and Montaldi (2015), who showed that the novelty effect reverses when the previously presented words are rare, i.e., it is the unexpectedness rather than relative familiarity that determines memory.

To understand further how PIMMS explains the relationship of novelty, surprise and memory, let us consider a level representing the current items perceived (e.g., objects), which may or may not be predicted from a higher level representing the context (e.g., environment). Take the example of sheep encountered in the urban environment: the context is predicting a typical urban configuration meaning seeing houses, road-signs, other cars, etc., whereas the sensory input is indicating the presence of sheep. This corresponds to an expectation that predicts only certain types of objects but the data indicates the presence of an unexpected object (Figure 1.2). This results in high PE, which causes strong encoding of the event. I refer to such situations as “surprise”, in keeping with other related work (Reichardt et al., 2020). In fact, this surprise could occur at multiple levels of the processing hierarchy. When it occurs between contextual predictions about familiar items, I call it “context surprise”. However, our knowledge of familiar items (e.g., sheep) allows us to make predictions about the perceptual features (e.g., four legs, white wool, etc.) that comprise those items. When one or more features differ from those expected (e.g., a pink sheep), then PE occurs at this lower level of the hierarchy. I call this “item surprise” (Figure 1.2B). Both types of surprise are similar to Berlyne’s (1960) concept of relative novelty. This type of surprise (and not novelty *per se*) is what drives memory encoding and leads to strong memory, however as I will show later that surprise (i.e. unexpectedness) is not the only way through which events can be encoded well into memory.

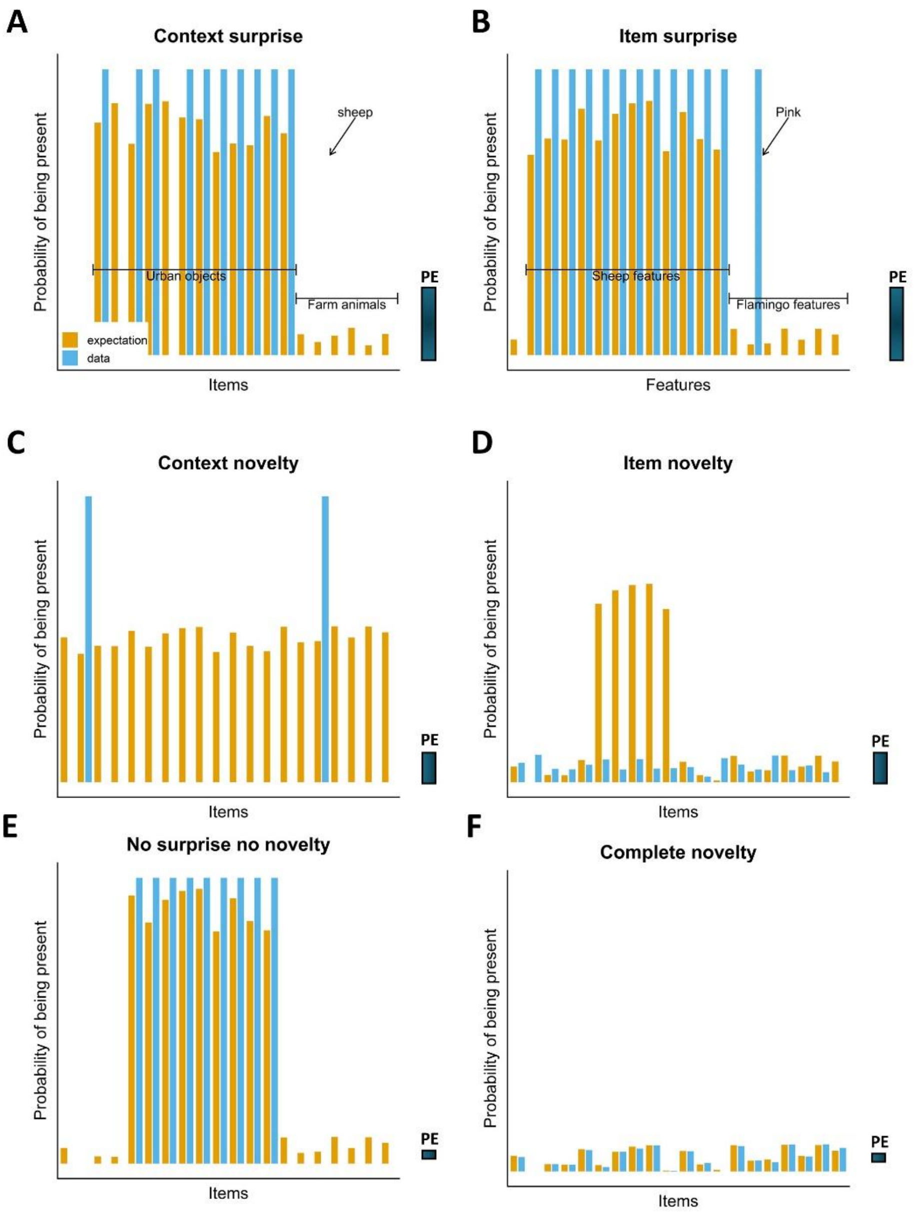

Figure 1.2: Schematic illustration of expectation and data over items and the likelihood of each individual item in the PIMMS framework: Items are assumed to be clear and unambiguous therefore the data is one if the items is present in the scene and zero if not. The bars reflect activity in a layer of neurons, where the discrete items or features on the x-axes are sorted arbitrarily by category labels - zoomed in merely for illustration. The x-axes only represent a tiny portion of large space of items or features so that right to the farm animals could for instance be sport equipment. The yellow bars represent the prior expectation from the "higher" level, whereas the blue bars indicate whether the item is known and is present in the scene, input from the level below (ultimately the sensory input). The data takes the value of one if present unless the semantic systems lacks representation as is the case for item novelty. The PE, which drives learning, is the difference between

expectation and data, whose magnitude is illustrated at the bottom right of each panel by a blue bar next to the main graphics. Panel A: precise expectation for items in a certain context, where one item is completely unexpected (context surprise). Panel B: the same as Panel A, except it refers to features of an object in lower perceptual level (item surprise). Panel C: Flat expectations and precise data (context novelty). Panel D: precise expectations and flat data (item novelty). Panel E: precise expectations and precise data with the for the same items (leading to no PE or learning; no surprise or novelty). Panel F: flat expectations and flat data, a combination of context novelty and item novelty (or "complete novelty"), but one predicted to show no PE or learning.

If you on the other hand consider a situation where you encounter an item you have never seen before, then the data is flat instead – i.e., you do not know how to interpret the sensory evidence. I call this "item novelty". This would correspond to the urban commuter encountering the sheep, but in this case, having no prior knowledge of animals like sheep at all.

In principle, where the expectation and the data overlap closely (Figure 1.2E), so that PE is low, there is no need to do any learning (i.e., no need to waste resources re-encoding what is already known). Interestingly, as I will show later, there are however situations where something is reasonably expected, yet it is still encoded well into memory. A low PE can also emerge from a flat expectations and flat data, as can be seen in Figure 1.2F. Importantly, this special case of "complete (or maximal) novelty" (e.g., encountering unknown objects in an unknown environment) is actually predicted to produce negligible, rather than "maximal", learning. In contrast to surprise, novelty in the sense discussed above is not associated with a memory benefit, which is at odds – at least on the terminological level – with several prominent models and accounts of novelty and memory (e.g. Lisman & Grace, 2005).

The key function of the contextual level is to optimise the predictability of an item occurring in a particular spatiotemporal context, i.e., store context-item associations. When we encounter a familiar item in a context that is different from what is expected, the ensuing PE induces learning of more accurate associations between episodic and semantic representations. This is the type of learning that enables memory of the spatiotemporal context in which an item occurred, or what has been called "recollection" (Mandler, 1980; Montaldi & Mayes, 2010; Yonelinas, 2002). This contextual level is associated with the hippocampus and other brain structures in the MTL and beyond (Aggleton & Brown, 1999; Mayes et al., 2007; Moscovitch, 1995; Moscovitch et al., 2016; Scoville & Milner, 1957; Squire, 1992).

Note that PEs can, and normally do, arise automatically, based on prior knowledge triggered by perceptual inputs, i.e., predictions are rarely intentional (Foster & Keane, 2015, 2019 for a different view). However, whereas Morris (2006) claimed that attended experiences are automatically recorded by the hippocampus, but normally fade and become lost, I claim that PE modulates the degree of encoding and therefore determines whether an experience will be available as a lasting memory. However, both views are not incompatible. As shown in Chapter 2, expected information per definition means low PE, yet there is ample evidence that schema-related information is encoded well, not despite of, but because of this information being expected. This may be possible because after automatic recording (Morris, 2006), prior knowledge helps to integrate expected information (see below for more discussion).

The semantic level on the other hand stores information about familiar items, and predicts which features are expected on the basis of a given item being present. There is a bidirectional flow of information between all levels, so not only do currently active item representations make predictions about associated perceptual features, but currently active features also influence which item representations remain active (i.e., perception involves dynamic competition across all systems in order to minimise overall PE). The semantic level is associated with anterior temporal lobe regions, including perirhinal cortex, and it is the strengthening of item-feature associations that enables the feeling of “familiarity” (rather than recollection), which sometimes accompanies recognition memory (Mandler, 1980; Montaldi & Mayes, 2010; Yonelinas, 2002), though some found that the hippocampus is also important for familiarity (Song et al., 2011).

Finally, while there are probably numerous levels of intermediate perceptual representations (as noted above), for example depending on the modality, for visual stimuli of the type used in this thesis, PIMMS associates posterior temporo-occipital regions associated with representing perceptual features, repeated processing of which can result in memory phenomena such as priming (Henson et al., 2003).

While the above section has illustrated how PIMMS can be applied to the literature of novelty and surprise and their role in declarative memory (for more extended discussion, see Quent, Henson & Greve, 2021), the aim of this thesis is to highlight and examine a number of shortcomings of PIMMS. The first shortcoming I address is that, while PIMMS explains how unexpected information is retained well, paradoxically the same seems to be true when information is highly expected, i.e., conforms to our prior knowledge or schema (e.g. Tse et al., 2007, 2011), addressed in the next section.

1.3 Expected as well as unexpected information is remembered well

Key to everyday functioning is the ability to predict aspects of our environment, and one source of such predictions are "schemas" (Bartlett, 1932). Schemas influence how we respond to and encode new experiences into memory (Bransford & Johnson, 1972), which in turn has the potential to update or form new schemas in the future to allow better predictions. Fernandez and Morris (2018) defined schemas as frameworks of knowledge (i.e. facts, skills but also attitudes) that are represented in a network of neurons that are connected and store memory traces. When these memory traces are activated, they alter responses to new information and therefore affect encoding, consolidation and retrieval. Clewett et al. (2019) simply views schemas as an "extracted rule based on knowledge". Others highlight in addition that schemas help us to generalise beyond experiences (e.g. Cockcroft et al., 2021). For the current purpose, schemas are active abstracted knowledge about recurring situations, such as what to expect when walking into a kitchen (compared to a bathroom, or other type of room), which affects cognition in general and memory in particular.

Numerous studies have shown that memory is better for information that fits with our prior knowledge, i.e., is expected from a schema (Alba & Hasher, 1983; J. R. Anderson, 1981; Craik & Tulving, 1975). This so-called "congruency effect" has been obtained using a broad range of memoranda (e.g. Bein et al., 2015; Brod & Shing, 2019; van Buuren et al., 2014). At the same time, many other studies show the apparent opposite finding, of better memory for unexpected or surprising information, i.e. that is incongruent with a schema (e.g. Brod et al., 2018; Greve et al., 2017; von Restorff, 1933; Worthen & Hunt, 2006), as is predicted by PIMMS.

This raises the question how information that is congruent or incongruent with a schema can both be remembered well. The neuroscientific model SLIMM (schema-linked interactions between medial prefrontal and medial temporal lobe; van Kesteren et al., 2012) proposes that memory is a U-shaped function of schema-congruency, i.e., the degree of expectancy of encountering a specific event, given a schema. More specifically, it is hypothesized that different brain systems support memory for the two ends of the U-shape: one based in the MTL and one based in neocortex, for which the medial prefrontal cortex (mPFC) plays a key gate-keeper role.

The MTL system (including the hippocampus) is assumed to encode highly unexpected (schema-incongruent) events, i.e., ones in which a schema's predictions are violated, producing PE. This system is therefore consistent with the PIMMS framework, though one subtle difference is that, whereas in PIMMS, PE drives local learning between the predictor and predicted, PE in SLIMM is assumed to trigger encoding of an episodic snapshot that captures all information present, regardless of whether it is relevant to the schema, i.e., including incidental contextual information ("global PE"). The evolutionary rationale for this is that, if something happens that is not predicted by the current schema, one wants to capture other contextual information that might be relevant to explaining why the schema failed, which (if it recurs) could be used to update (or create a new) schema in future. For example, if one is surprised to find boiling water coming from a tap on a kitchen sink, one should remember details about that tap, and update one's kitchen schema so that, when encountering such taps again, one knows that making tea can be done using boiling water from these taps (rather than from a kettle).

The other neocortical system involves the medial prefrontal cortex (mPFC), which best encodes highly expected (schema-congruent) events. SLIMM assumes that, when high "resonance" occurs between a schema and the perceived environment, the mPFC inhibits the MTL system (preventing an episodic snapshot) and instead augments direct encoding of the event into neocortex (even if that encoding is transient). Importantly, only information that is relevant to the schema is encoded, such as the location of a kettle used to make tea in a new kitchen that one enters (given that making beverages is one function of a kitchen), whereas incidental information, such as the colour of the kettle, is lost.

Evidence for these two systems comes from Tse and colleagues (Tse et al., 2007, 2011). These authors showed that schemas allowed animals that were trained to remember flavour-location pairings to quickly assimilate new information, which then became independent from the hippocampus faster than would be expected by the standard systems consolidation view. Importantly, this type of fast learning was accompanied by up-regulation of immediate early genes in mPFC and pharmacological intervention targeting mPFC abolished fast acquisition and retrieval of flour-location pairings. This suggests that the long route of system consolidation (from hippocampus to neocortex) can be bypassed via the mPFC if a schema exists.

In humans, Greve et al. (2019) provided the first behavioural evidence for the U-shape predicted by SLIMM. The authors used a paradigm in which the schema was an abstract rule about the relative value of two types of objects, and episodic memory was tested for trials containing unique numbers of each object type. More specifically, these authors tested episodic memory under three levels of expectancy/congruency: a congruent condition in which a trial was as predicted by a rule, an "unrelated" condition where there was no rule (or only a weak rule), and an incongruent condition in which a trial broke a strong rule. Across four experiments, memory for unique trials was worse in the unrelated (middle) condition than in the two others, i.e., there was a U-shaped function across the three conditions. Furthermore, memory for the two extreme conditions (congruent and incongruent) appeared dissociable by other variables, as is predicted by SLIMM if they are supported by different memory systems with different characteristics. For example, memory for incidental information was only improved in the unrelated condition, as predicted by SLIMM when a PE triggers encoding of an episodic snapshot.

Nonetheless, there are several limitations to the Greve et al. (2019) study. Firstly, with only three levels of schema-congruence, it is possible that memory was worse in one condition than the other two (particularly the middle one) because of differences between conditions other than expectancy (such as level of task engagement when there is no clear rule, which the authors tried to address by various manipulations, but these may not have been sufficient). A more elegant approach would be to manipulate expectancy in a continuous fashion (i.e., parametrically rather than categorically). Secondly, the "schema", while an abstract rule, was arguably a trivial case of the much richer schemas that operate in everyday life. Thirdly, that schema had to be learned across several "training" trials, which made the experiment long, and may also have produced different levels of schema strength across participants (even if on average their behaviour showed good learning of the rule). Furthermore, while possibly adding a level of experimental control, any paradigm in which schemas are learned during the experiment would be difficult to apply to patients (such as those with MTL lesions), who may not be able to learn a new schema in the first place (i.e., any decrement in their memory for trials relative to controls could reflect poor schema learning rather than impaired episodic encoding). Many of these problems can be overcome by using pre-experimental schema, e.g., everyday schema that are learned over many years prior to participating in a laboratory experiment (or prior to a brain lesion). Therefore, I developed a paradigm that uses pre-experimental knowledge about kitchens (and what one expects to encounter in them).

The details are described in Chapter 2, but in brief, the aim was to replicate a U-shaped function of memory for new information as a more continuous function of schema-congruence (or what I will call “expectancy”), using a much more realistic, pre-experimental schema. To this end, I used immersive virtual reality (iVR) to “place” participants in a life-like situation (a virtual kitchen), and tested their memory for the locations of objects to examine it as a function of how likely those objects were to appear in that location (i.e., based on their pre-existing schema for typical locations to find objects in a kitchen). So for example, a kettle might appear on the kitchen counter (expected), on a kitchen table (neither strongly expected nor unexpected) or on top of a trashcan (unexpected). I then tested participants’ memory using recall, where they had to place a given object at the location where they remembered it, and recognition, where they had to select which of the three images of an object in different locations corresponded to the location they remembered (i.e., three-alternative forced choice, 3AFC).

One possible explanation for the U-shape is that unexpected locations are surprising, attract attention and this improves memory encoding. Expected locations, however, are not necessarily encoded better, but simply benefit at retrieval owing to the tendency for participants to guess an expected location (based on their schema) when they cannot remember the specific location. To control for the latter guessing effect, I ensured that the two foils used in the 3AFC test showed locations for the target object that were approximately equally expected (based on normative ratings). If the U-shape remains in 3AFC performance, then this suggests that the memory advantage for expected locations also occurs at encoding, supporting the idea of two separate memory systems.

A second prediction of the SLIMM model follows from its assumption that the two memory systems operate according to different mechanisms, evolved for different goals. As noted above, it makes sense for the MTL system that underlies the memory advantage for surprising events (those that produce a high PE) to encode an episodic “snapshot” of such events, which includes surrounding contextual information that appears incidental to the prediction (schema). This is different to PIMMS, which assumes only associations relevant to the schema will be strengthened by PE, not incidental details. On the other hand, the mPFC system underlying memory for expected information is assumed to enable rapid integration of new information into an existing schema, during which incidental (contextual) information is lost, because it is not part of that schema (van Kesteren et al., 2012). SLIMM therefore predicts that memory for expected information will be associated with a feeling of familiarity instead, in the absence of recollection of the original context. I therefore used a variant of the “remember/know” procedure (Tulving, 1985) to assess whether memory for unexpected locations was more likely to be associated with “remember” responses while memory for expected locations was more likely to be associated with “familiar” responses.

In summary, Chapter 2 is centred on the question that PIMMS cannot predict this U-shape, while SLIMM can. A second limitation of PIMMS addressed in this thesis relates to the fact that it is silent on neurobiology, and therefore cannot explain some phenomena related to PE and memory that appear to relate to the processes of cellular/synaptic consolidation.

1.4 PIMMS is silent on neurobiology

In Chapter 3, I turn to the question how experiencing something novel influences our memory for information that occurs in close temporal proximity to that experience? As noted above, it is well established in both human and non-human animals that novel stimuli that are surprising are remembered better than familiar stimuli (e.g., Ranganath & Rainer, 2003; Tulving & Kroll, 1995; van Kesteren et al., 2012). A neurobiological explanation for this is the increased neuromodulatory influences of acetylcholinergic and noradrenergic systems (Ranganath & Rainer, 2003) and/or of dopaminergic regions in the midbrain (Bunzeck & Düzel, 2006; Lisman et al., 2011; Lisman & Grace, 2005). According to one framework for example (Lisman et al., 2011; Lisman & Grace, 2005), novelty is detected in the hippocampus, which sends a signal to the ventral tegmental area, leading to release of dopamine in the hippocampus via dopaminergic back projections, which in turn lowers the threshold for learning (Schomaker, 2019). While novelty and surprise are not necessarily identical (Quent et al., 2021), these neurobiological explanations are not part of PIMMS.

Interestingly, experiencing something novel can not only enhance memory for the novel information itself, but also affect memory for other, unrelated information that occurs in temporal proximity to the novel information (Fernández & Morris, 2018). This enhancement of memory for information occurring either before or after the novel experience has been shown in both non-humans (Ballarini et al., 2009; Moncada & Viola,

2007) and humans (Ballarini et al., 2013; Fenker et al., 2008; Ramirez Butavand et al., 2020). Novelty-related memory enhancement has been found in a variety of paradigms: inhibitory avoidance (Moncada & Viola, 2007), spatial memory (Wang et al., 2010), spatial object recognition (Ballarini et al., 2009), contextual fear conditioning (Ballarini et al., 2009) conditioned taste aversion (Ballarini et al., 2009), story and picture recall (Ballarini et al., 2013) and word learning (Fenker et al., 2008; Schomaker et al., 2014). In possibly the most real-world application of this novelty effect in humans, Ballarini et al. (2013) showed that memory of primary school children was enhanced if they attended special science or music lessons that were designed to be novel. The enhancing effect was observed for the learning of other, unrelated verbal and pictorial information, provided the novel lesson took place within an hour before or after learning, consistent with a critical time window during which the novelty effect operates (see below).

One neurobiological explanation for the effect of novelty on surrounding information is provided by "behavioural tagging theory" (BTT; Ballarini et al., 2009; Moncada & Viola, 2007), which itself derives from the physiological mechanisms proposed by the synaptic "tag-and-capture" theory (Frey & Morris, 1997; Redondo & Morris, 2011). Briefly put, this theory postulates that two main steps are important to maintain late-long-term potentiation (late-LTP). First, a synapse is tagged because it has received input. In the second step, the tagged synapse needs to capture so-called plasticity-related products (PRP) in order to induce the lasting structural changes that give rise to late-LTP. Experimentally it can be shown that strong tetanisation of a synaptic input can produce both tagging and subsequent PRP capture. Weak tetanisation of a synaptic input, on the other hand, induces early-LTP, but this is not maintained unless the second stage of PRP capture occurs. One way to produce this PRP capture is to provide a second, strong tetanisation to a different synaptic input on the same population of neurons. In that case, both synaptic inputs benefit from the provision of PRPs and hence late-LTP is maintained. Something similar to strong tetanisation can potentially come from a different, but highly novel input, leading to a similar maintenance of late-LTP (Li et al., 2003; Straube, Korz, & Frey, 2003; Straube, Korz, Balschun, et al., 2003). Both the induction of LTP in the hippocampus and behavioural tagging are dopamine dependent (Li et al., 2003; Wang et al., 2010), consistent with the aforementioned idea that dopamine is crucial for novelty-related memory enhancement (Lisman et al., 2011; Lisman & Grace, 2005).

Within the "tag-and-capture" theory, the lifetime of a tag is limited to approximately 90 minutes (Redondo & Morris, 2011), requiring the weakly-learned information and strong tetanisation to co-occur within that time window. Likewise, behavioural tagging only occurs within a certain time window (Ballarini et al., 2009; Moncada & Viola, 2007). For instance, weak inhibitory avoidance training that normally only leads to spatial STM can be strengthened to LTM if animals are allowed to explore a novel, open field within up to an hour of that training (Moncada & Viola, 2007). However, the timescale of the behavioural tagging window depends on the task characteristics and may even have a nonlinear expression, given that some animal studies have shown that novelty that is too close to the unrelated encoding event does not enhance memory (Moncada et al., 2015). Information about the temporal dependencies in humans is scarce however, and several studies have shown memory enhancement when a novel experience occurs within a few seconds of the learning experience (e.g. Bunzeck & Düzel, 2006; Schomaker et al., 2014).

In addition to the time between the encoding of critical information and the novel experience, a second consideration is the time between the novel experience and the subsequent test of memory (retention interval). Most animal models assume that a period of consolidation is required, such that the effects of novelty only emerge after a delay. However, in humans, Bunzeck and Düzel (2006) showed that presenting familiar images at the same time as novel images led to an overall increase in memory performance for the familiar images in a subsequent recognition memory task, but only when recognition was tested immediately; not when tested the next day (see also Biel & Bunzeck, 2019, for a failure to find an effect of novelty the next day). The effect of retention interval and possible role of consolidation therefore remains unclear.

According to BTT, there are at least two further boundary conditions for behavioural tagging. Firstly, as noted above, novelty does not enhance memories traces that are already strong, presumably because they already sufficiently captured PRPs (Moncada & Viola, 2007). This may explain why the effect of novel lessons on children in the above Ballarini et al. (2013) study was most pronounced for difficult information, which presumably would have only led to weak memories otherwise. It is also consistent with recent findings related to stress. Like novelty-related memory enhancement, stress-related memory enhancement have been linked to processes akin to those hypothesized in tag-and-capture theory (Bergado et al., 2011; McIntyre et al., 2012; Richter-Levin & Akirav, 2003). For example, spatial recognition memory in rats was promoted from STM to LTM by acute stress after weak but not after strong training (Lopes da Cunha et al., 2019), and stress-related increases of cortisol in humans only predicted memory for weakly-learned neutral words, but not for strongly-learned reward-predicting words (Quent, McCullough, Sazma, Wolf, & Yonelinas, 2018; see Dunsmoor, Murty, Davachi, & Phelps, 2015, for similar effects using Pavlovian fear conditioning).

However, the suggestion that novelty preferentially aids weak memories is less easy to reconcile with other human studies. For example, Fenker et al. (2008) used a paradigm similar to Bunzeck and Düzel (2006) to demonstrate that presenting novel images enhanced memory for unrelated words, but they only found this enhancement for words whose recognition was accompanied by a "Remember" judgment; the advantage was not seen for words whose recognition was accompanied by a "Know" judgment. Remember and Know judgments are associated with the theoretical concepts of recollection and familiarity, and while these concepts are not synonymous with memory strength (Yonelinas, 2002), few would contest that items judged familiar have, on average, weaker memory representations than those recollected. It is possible then that some words in Fenker et al.'s study were initially encoded weakly, but the novel experience boosted them sufficiently that they were later recollected. However, it also seems likely that some words could have be encoded so weakly that they would not be recognised at all (i.e., missed), had a novel experience not boosted them such that they at least seemed familiar. In this case, an effect of novelty would be expected on Know judgments as well as Remember judgments, and possibly more so, if Know judgments are a better indicator of items that were initially encoded weakly.

One possible explanation for Fenker et al.'s finding that novelty improved recollection relates to a second boundary condition of BTT: that information must converge on the same neural population that is activated by the novel experience, in order for that information to benefit from the novelty-induced PRPs (Ballarini et al., 2009; Nomoto et al., 2016). For instance, open field exploration does not enhance conditioned taste aversion in rats, but a novel taste does (Ballarini et al., 2009). Here it is important to distinguish conceptually-unrelated (e.g. word learning and spatial navigation) and neuronally-unrelated. In its most basic form, BTT postulates that detection of novelty leads to a dopaminergic signal that boosts encoding in places that receive that signal. It is therefore conceivable that novelty is detected by one neural population within a brain region (e.g., the hippocampus), but a wide-spread signal is also received by other populations within that region, including the population that is encoding task-relevant information. Therefore, information that is conceptually-unrelated to the novel experience can still benefit providing the information is neuronally-related in the sense of receiving the same memory-boosting signal. Given the above evidence that hippocampus is important for detecting novelty, and other evidence that the hippocampus is important for encoding the spatiotemporal and associative context that defines recollected memories, then it is possible that a novel experience only improves recollection of information (as in Fenker et al., 2008).

A further consideration is the nature of the novel experience. Previous human studies have used novel images or films, and at least one of these (Biel & Bunzeck, 2019) recently failed to find an effect of novelty. This study compared the effects of watching novel versus familiar films, and the authors speculated that the lack of difference was because the films did not engender active engagement, at least to the level engendered by the exploration of a novel spatial environment that is used in many animal studies. Furthermore, if the hippocampus is key for the novelty effect, active navigation might be important for maximally engaging the hippocampus, given its role in navigation (Burgess et al., 2002; O'Keefe & Nadel, 1978). One way to expose humans to a novel environment, but within the controlled setting of a laboratory, is to use VR. Indeed, another human study used VR to compare novel versus familiar environments in their effects on words learned immediately after the experience (Schomaker et al., 2014). Exploring a novel environment, relative to a familiar one, enhanced free recall of the words, though not recognition memory for the words (though these authors did not distinguish recollection versus familiarity in their recognition task). Since recall requires recollection, these findings are consistent with Fenker et al. (2008), on the assumption that their recognition performance was dominated by familiarity.

Given the importance of these findings for education and other real-world situations, in Chapter 3 I attempted to replicate the effects of a novel spatial navigation experience on memory for unrelated words, presented immediately prior to the novel VR experience, as a function of the encoding task (deep vs. shallow) and retrieval quality (recognition with remember vs. know judgments, plus recall), to test whether weakly learned information and recollection would be most affected by the novel experience.

1.5 PIMMS fails to explain the effects of event boundaries

A final limitation of PIMMS I consider is the effect of "event boundaries". While PIMMS describes dynamic interactions between representational layers, it treats the information that is encoded as discrete experiences, without any consideration of how we segment our continuous experience into such discrete chunks or "events". Indeed, it has been proposed that PEs, of the type considered by PIMMS, are the way that we segment experience, by creating “event boundaries” whenever something happens next that we do not predict, based on our current schema (Zacks et al., 2007). These boundaries may then be points in time at which we “bind” the information in STM/WM from the preceding event, in order to create a LTM trace (Ben-Yakov & Henson, 2018).

While PEs are one possible trigger for event boundaries, according to the event segmentation theory (EST; Reynolds et al., 2007; Zacks et al., 2007), a recent review by Clewett and colleagues (2019) suggests that general shifts in space, goal or perception also lead to event boundaries. Moreover, these authors reviewed evidence that experiencing a boundary can impair memory for the temporal order of information, as well as increased estimates of temporal distances and overestimation of temporal durations. For example, many studies using simple sequences of pictures on different backgrounds report superior memory for the order of two pictures when they occur within same event (background) compared to when they straddle an event boundary (different backgrounds) (DuBrow & Davachi, 2013, 2016; Ezzyat & Davachi, 2014; Heusser et al., 2018). At the same time, item recognition and item-context binding can be enhanced by boundaries (Heusser et al., 2018; Rouhani et al., 2020; Swallow et al., 2009).

A more typical, everyday experience of event boundaries occurs when passing through doorways between rooms. For example, when you get up from the couch in the living room to go to the kitchen, you sometimes forget what you wanted to do as soon as you walk through the door. This effect is called the *location updating* or *doorway effect,* and has been extensively studied (Lawrence & Peterson, 2016; McFadyen et al., 2021; Pettijohn & Radvansky, 2018a, 2018b; Radvansky et al., 2010, 2011; Radvansky & Copeland, 2006; Seel et al., 2019). For example, Radvansky (2006) asked participants to navigate through a series of rooms that contained objects on top of tables. Participants picked up an object from one table and carried it to the next one, where they put the object down and picked up another one. Periodically, memory was probed for the objects that were either just put down or picked up, and memory was typically lower if the participants had just walked through a doorway. To account for the location updating effect, the EST has been further developed to the Event Horizon Model (EHM; see Radvansky et al., 2011). This model posits that: 1) events are created by boundaries for which event models are generated, but people can only process one event model at a time; 2) the current event model foregrounds information pertaining to that model; 3) retrieval is facilitated for non-competitive retrieval but impaired for competitive retrieval and 4) causal relations among events are stored. The third property explains why walking through a door impairs item memory: walking through a door creates a new event model, but the most recent object is associated with the previous room (event model), and so memory is impaired as both event models compete for retrieval (though see Logie & Donaldson, 2021, for alternative accounts, as discussed further in Chapter 5). Horner et al (2016) found that doors between rooms (in virtual reality) also affect memory for the temporal order of objects encountered within versus between those rooms.

However, it remains unclear what aspect of walking through a doorway renders it a boundary. It could be that there is something special about doors, but it could also be that context shifts that result from going from the living room to the kitchen are responsible. However even if the encoding context was reinstated (by returning to the room), the location updating effect can still be observed (Radvansky et al., 2011). On the other hand, walking through a door also typically leads to perceptual change (PC). For instance, you probably have different kinds of objects and furniture in the kitchen than you have in the living room, as a consequence you experience PC when you walk through that door. The obvious way to tease these apart is to have a PC in the absence of a door, and/or a door in the absence of a PC. Or perhaps neither PC or a door matter, and what really matters is a PE. Other possible reasons that I did not plan to examine with this paradigm but are potentially relevant are context and schema shifts. These were my general aims in Chapter 4, i.e. to try to tease apart the effects of doors, PC and PE on the formation of event boundaries, as measured by their effects on memory for temporal order that objects were encountered within versus across rooms in a virtual environment, similar to Horner et al. (2016); though in the end, I could not replicate a basic effect on temporal order memory in my version of this paradigm, as discussed later in Chapter 4.

1.6 A note on the statistics used in this thesis

Before moving to the empirical parts of this thesis, it is important to consider the statistical approach I took. There is a recent and growing interest in using Bayesian inference, rather than the more typical frequentist approach (e.g., p-values), for scientific claims (Dienes, 2016; Morey et al., 2016; Wagenmakers et al., 2011). One reason is to be able to quantify evidence against an effect, which can be expressed through the Bayes Factor (BF) for one hypothesis (e.g., the alternate hypothesis H_1) relative to another (e.g., the null hypothesis H_0):

$$\frac{p(H_1|\text{data})}{p(H_0|\text{data})} = \frac{p(H_1)}{p(H_0)} \times \frac{p(\text{data}|H_1)}{p(\text{data}|H_0)}$$

$$BF_{10} = \frac{p(data|H_1)}{p(data|H_0)} = 1/BF_{01}$$

In frequentist statistics, one cannot easily quantify evidence against H_0, but rather only report the probability (p-value) of H_0 producing the data (i.e., a statistic as high or higher), and so if this is not small enough (e.g. $p > .05$), one is left with absence of evidence rather than evidence of absence.

Another reason for a Bayesian approach is to facilitate more efficient experimental designs, called "sequential designs" (Schönbrodt & Wagenmakers, 2018), where data can stop being collected as soon as a BF (either in favour of the null or in favour of the alternative) surpasses some criterion (see table below). This is more problematic for frequentist approaches (Dienes, 2016). I used sequential designs in Chapters 2 and 4.

While the scientific field has settled on a conventional frequentist p-value (alpha) of .05 for declaring something as "significant", less consensus has been reached about what BF constitutes "strong evidence" for example. While some Bayesians would prefer not to categorise BFs with verbal labels, it is helpful to assign some verbal labels to ranges of BFs (Jeffreys, 1998; Kass & Raftery, 1995). For this thesis, I have adopted what is currently regarded as "publishable" evidence by various journals in our field, which typically either accept $BF_{10} > 6$ and $BF_{10} < 1/6$, or $BF_{10} > 10$ and $BF_{10} < 1/10$. I therefore adapted the previous conventions (Jeffreys, 1998; Kass & Raftery, 1995) as shown in Table 1.1.

Table 1.1: Convention how to label evidence in terms of BF.

BF_{10}	EVIDENCE
> 100	Extreme evidence for H_1
30 – 100	Very strong evidence for H_1
10 – 30	Strong evidence for H_1
6 – 10	Moderate evidence for H_1
3 – 6	Anecdotal evidence for H_1

3 – 1/3	Inconclusive evidence
1/3 – 1/6	Anecdotal evidence for H_0
1/6 – 1/10	Moderate evidence for H_0
1/10 – 1/30	Strong evidence for H_0
1/30 – 1/100	Very strong evidence for H_0
< 1/100	Extreme evidence for H_0

Finally, there is always debate about the appropriate priors to use for Bayesian inference, since the evidence (BF) depends on the priors for each parameter in the statistical model. This is an advantage of pre-registering analyses, which protect one from being accused of altering one's priors post hoc (after seeing the data), and most of the analyses in this thesis were (pre-)registered on the Open Science Framework (OSF, https://osf.io; with the exception of some of the initial, pilot experiments in Chapters 2 and 4, and when problems with original analyses, or useful additional analyses, became apparent between one experiment and the next, as in Chapter 2). Here I used standard recommendations for those priors, which depend on the type of statistical model used (e.g., T-tests, ANOVAs or logistic regression). These are "objective" priors that enable estimation of the probability that the data came from H_0, versus probability that the data did not come from H_0 (i.e., H_1 = ~H_0, rather than "subjective" priors for H1, namely that the data came from a different distribution with a specific effect size). For some of my experiments, I also simulated the Bayesian equivalent of "power", i.e., the probability of obtaining a BF above a certain threshold as a function of the sample size and the effect size (by generating random data with an effect size of 0 for H_0, or, say, a Cohen's d of 0.5 for H_1).

1.7 Summary

To sum up, I have outlined three prominent areas that are related to PE, novelty and memory, but which cannot currently be explained by PIMMS. First, PIMMS cannot explain why highly expected information is encoded well (Chapter 2); second, PIMMS cannot explain why memory for information surrounding a novel experience is often also remembered well (Chapter 3); and thirdly, PIMMS cannot explain how continuous stimuli are segmented by event boundaries and how those boundaries affect memory (Chapter 4). In brief, Chapter 2 confirmed a U-shaped function of memory against expectancy, for which the advantage for expected events is not predicted by PIMMS, but is consistent with the related SLIMM model, while Chapters 3 and 4 found no evidence for (and sometimes evidence against) an effect of novelty on preceding information or an effect of doorways on temporal order memory, respectively. Chapter 5 discusses all these findings, in particular the extent to which the results in Chapters 3 and 4 reflect limitations of the paradigms used, versus remain fundamental problems for PIMMS.

2 THE EXPECTED CASE

2.1 Introduction

As explained in Chapter 1, PIMMS is unable to explain how expected information is remembered well because its lack of PE. So to show both expected (schema-congruent) as well as unexpected (schema-incongruent) information is encoded well, I used iVR to test people's memory for the location of everyday objects within a virtual kitchen. The expectancy of each object-location was derived from normative ratings of 20 objects in each of 20 possible locations, from which pairings were selected to cover a range from highly expected (+100) to highly unexpected (-100; see below for more information). Participants spent only 45 seconds exploring the virtual kitchen and were instructed to count these 20 objects, before the objects were removed for a short delay, and participants were asked to place individual objects at their remembered location, followed by the final 3AFC test outside iVR (using stills on a computer screen).

This paradigm was inspired by a study (Lew & Howe, 2017) that used realistic photos of schema-evoking rooms (bathroom, kitchen, living room and office) to investigate how expectedness of locations for specific objects affected memory for those objects and for their locations. The photos showed room-congruent objects at expected and unexpected locations, as well as room-incongruent objects. At test, objects either stayed in same location or shifted to a different location. Recognition memory for objects was better for room-incongruent objects as well as room-congruent objects at unexpected locations, relative to room-congruent objects at expected locations. However, when memory was tested instead by recall of an object's location, memory for objects at unexpected locations was impaired relative to room-congruent objects at expected locations. The authors explained their findings in terms of schemas acting differently on item and associative (location) memory, whereby an unexpected location attracts attention, but also activates schema-congruent bindings that interfere with memory (see also Bower et al., 1979).

However, another possible explanation for Lew and Howe's finding of impaired recall of the location of objects at unexpected locations is that, if participants forgot the location, then they guessed based on a schema. For example, if you forget where you saw the saucepan, then you can use prior knowledge to guess that it was on the stove. A similar guessing bias has been used to explain the advantage of schema-congruent information in source memory (Bayen et al., 2000; Kleider et al., 2008; though also see van der Linden et al., 2017). This guessing bias could have obscured any advantage in memory for the location of objects at unexpected locations. One way to address this bias (as used here) is to test memory with forced-choice tests, in which the target (saucepan on stove) and foil (saucepan in cupboard) are matched on their expectancy. Furthermore, Lew and Howe only had two levels of expectancy, so could not test for a U-shaped function of memory that the SLIMM model predicts (see Chapter 1). I therefore used a continuous (parametric) range of expectancies, based on the participants' ratings.

My two main predictions were: 1) recall and recognition for object-location would be a U-shaped function of expectancy of an object's location, as tested by a quadratic component and interrupted linear regression, and 2) recollection, as estimated from remember/familiar judgments, would decline with expectancy, while familiarity would increase with expectancy. To allow for individual differences in schema, expectancy ratings were defined individually, based on a debriefing phase.

I tested these two hypotheses across four experiments. The first experiment was an initial attempt to obtain a basic U-shape, while the next three iterations added a measure of recollection/familiarity to additionally test the second hypothesis. Apart from the first experiment, the experiments were preregistered, and I used Bayesian analysis to propagate evidence across experiments, i.e., using the posterior of one experiment as the prior for the next experiment. First, a small normative study was run to determine at which pre-defined locations to place the objects, in order to allow sampling from the whole range of expectancy (from very unexpected to very expected).

2.2 Normative study

In order to select where to place each object, and to get a range of expectancy values, six participants were shown screenshots of each object at each location and asked to rate how expected that object was in that location. They were then asked to also rate the general expectancy of each object in a kitchen.

Methods

Participants

Participants were members of the research group (five females and one male; age M = 36.83 (SD = 2.14)).

Materials

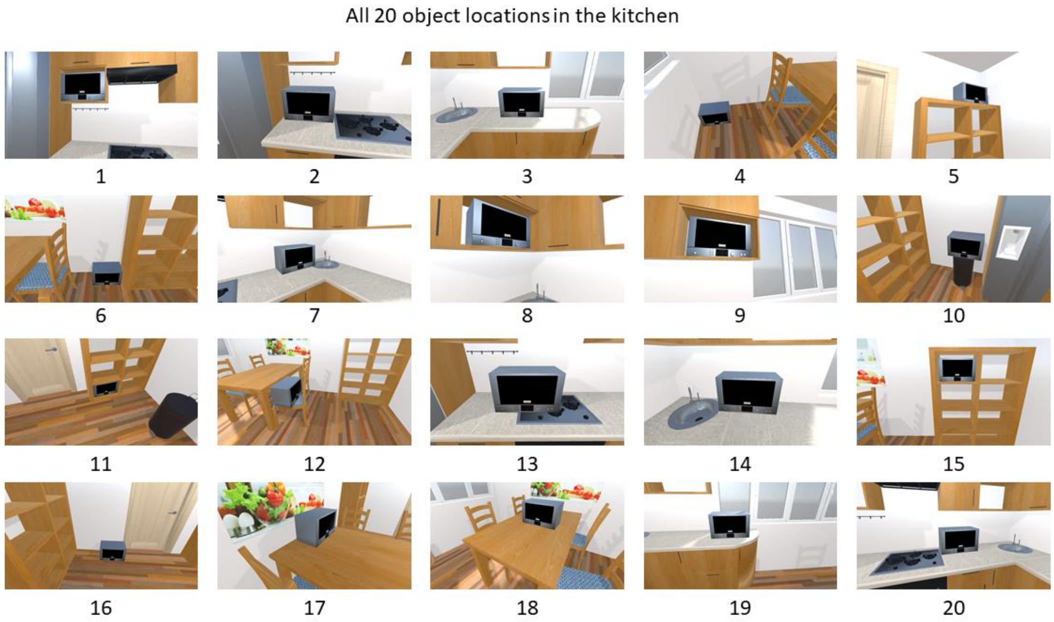

Figure 2.1: All object locations in the kitchen: Pre-defined object locations are shown with microwave as the example object.

The virtual kitchen was 5.15 by 4.40 virtual-metre (vm) kitchen, where 1 vm corresponds to approximately 1 metre in the real-world (material available at https://osf.io/4sw2t/ and https://github.com/JAQuent/schemaVR). 20 objects were placed at 20 locations within the kitchen, and then a still image captured of each object in each location (see Figure 2.1). The objects (Table 2.1) consisted of 12 kitchen objects and 8 non-kitchen objects, inspired by Lew & Howe (2017). Within the kitchen, 20 places were defined as canonical object locations where all objects could be placed. Due to the different sizes, which ranged from a large microwave to a small kitchen knife, not every place in the kitchen was suitable to become an object location. While some of the pre-defined locations where selected with certain objects in mind (e.g. the stove top was chosen with the sauce pan in mind) to include very expected locations, some were chosen for the opposite reason (i.e. being very unexpected locations like in the corner of the floor). Generally, object locations were spaced out, however – as in a normal kitchen – certain areas like the working surfaces of the kitchen contained a higher density of object locations.

Table 2.1: List of kitchen objects and non-kitchen objects used: Table shows the object names and their numbers but not the locations. This can be seen in the Appendix

Kitchen Objects				**Non-Kitchen Objects**	
1 microwave	5 bowl of fruits	9 bread		13 towels	17 hat
2 kitchen roll	6 tea pot	10 glass jug		14 toy	18 helmet
3 saucepan	7 knife	11 mug		15 books	19 calendar
4 toaster	8 mixer	12 dishes		16 umbrella	20 fan

Participants rated how likely each of the 20 objects was to be found in each of the 20 locations (i.e., 400 trials in total) from -100 to +100 using a slider. Four additional objects (kitchen: peppers and a white pot; non-kitchen: a dumbbell and wrench) were used to create eight object/location practice trials, which were shown first to give participants an idea about the task and calibrate their ratings. Results are shown in Figure 2.2.

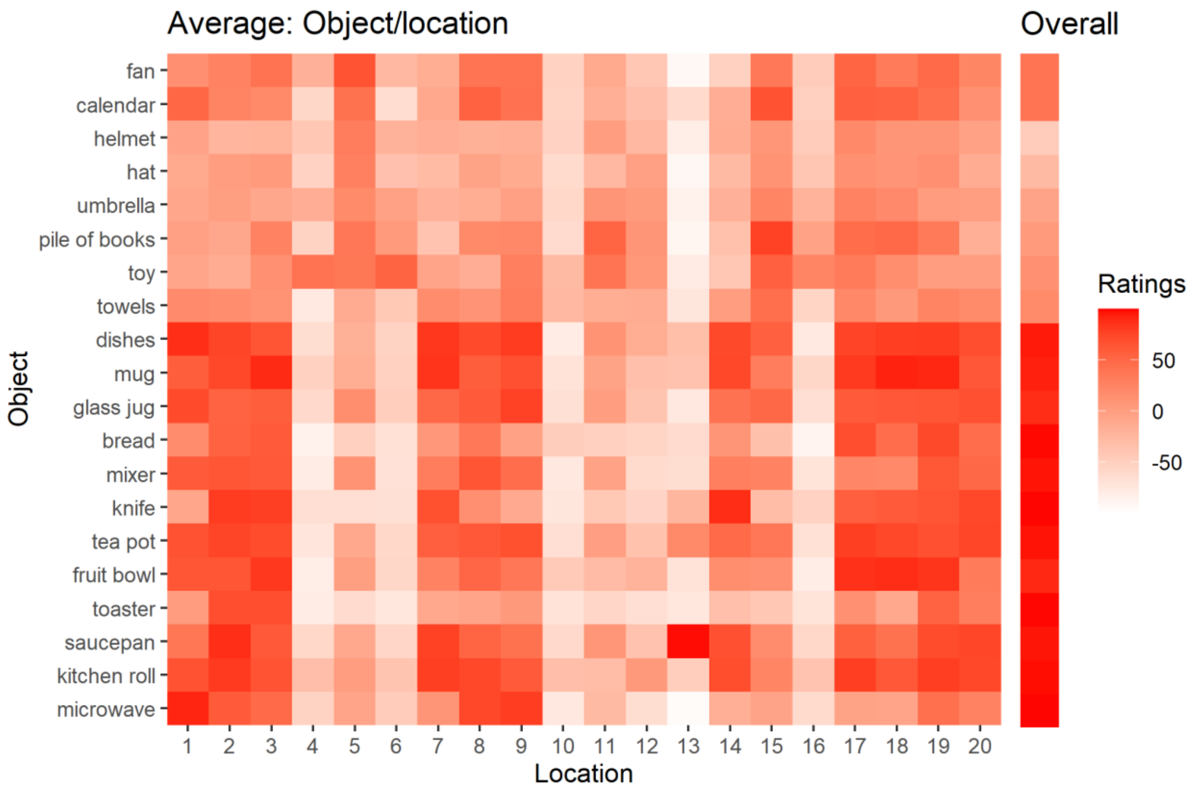

Figure 2.2: Mean normative rating of object-location for each object at each location: Individual ratings could vary from -100 to 100. The column titled 'overall' (right) shows average general expectancy rating for each object.

These ratings were then ranked from 1 and 400 within each participant (since different participants used different ranges of values), and the ranks were averaged over participants. Then an algorithm was run with close to 10 million iterations in which each object was randomly assigned to the 20 locations along with two foil locations that were randomly chosen (the foil locations were used in the 3AFC test of memory, and should have expectancies close to that of the target location). When the solution was not valid (e.g. the same location used for foil 1 and foil 2), this step was repeated until a valid solution was returned. For each valid solution, the sum of squared differences between the ranks of the targets and the intended uniform spread ("SS of targets") and the sum of squared differences between the ranks of the targets and both foils ("SS of foils") were calculated. In a two-step process, the number of solutions was first reduced to only include the 0.0001th percentile of the "SS of foils" distribution, to prioritize adequate foils that are similar in expectancy to the target. Second, the remaining solutions were sorted by their "SS of targets" values and then checked for problems (e.g. some object not being visible). The first valid set of 20 object-location pairs that was found by this algorithm is shown in the Appendix (see Table 7.2).

2.3 Experiment 1

Experiment 1 used the optimised set of 20 object-location pairs from the normative experiment to test for a basic U-shape for object-location memory as a function of expectancy of that location, and estimate its effect size. This was defined by the quadratic coefficient of a second-order polynomial expansion of individuals' continuous expectancy ratings.

Methods

Participants

All participants in the study were Cambridge community members from the volunteer panel of the MRC Cognition and Brain Science Unit or recruited through word-to-mouth from the wider Cambridge community, all of whom had reported normal or corrected-to-normal visual acuity, provided informed consent and received monetary compensation for participation, as approved by a local ethics committee (CPREC 2020.018).

Because Experiment 1 was the first experiment on this paradigm, no effect size was available to power the study, and so I decided to test 16 participants, as a number typical of laboratory memory experiments. These participants had a mean age of 26.38 (SD= 3.52), with eight females and eight males. For one participant, information about whether the objects were seen or not at study was lost. However, recordings of their object placement was intact, so I kept this participant, given that it was rare for objects to be reported as not seen.

Materials

The first set of 20 object-location pairings that were optimised from the normative study (Set 1) were used.

Procedure

The basic paradigm for all four experiments is illustrated in Figure 2.3, and contains an encoding and recall phase performed in iVR, followed by a recognition and expectancy-rating phase performed on a computer screen. During encoding, participants were asked to navigate freely through a virtual kitchen for 45 sec, with the instruction to count and memorise the locations of all 20 objects that were scattered across the room. A small amount of exploration (i.e., movement and rotation) was required to detect all 20 objects. Following encoding and prior to recall, participants entered a blank environment for approximately 2.5 min to practice how to place objects (simple cubes) using the iVR hand controls; a skill that was needed in the subsequent recall phase. For recall, participants then re-entered the kitchen (now without the objects being present), were given one object and asked to place it at its previously seen location (once placed, the object disappeared, and the process was repeated for the remaining 19 objects). Participants were encouraged to guess if they were unsure, but could skip if they did not remember the object at all (a miss). *Recall accuracy* was defined as whether or not the correct location was the closest of the 20 object locations to the recalled location.

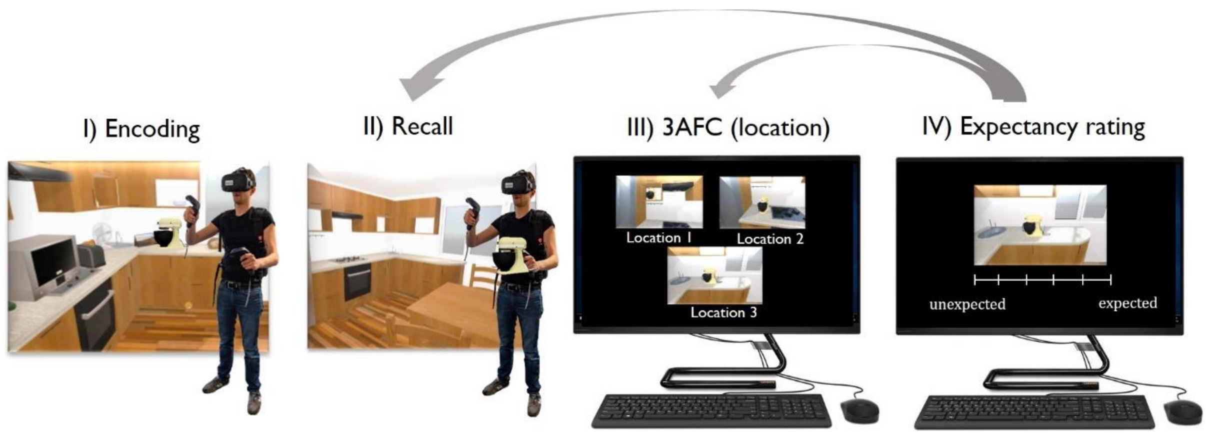

Figure 2.3: Schematic overview of paradigm: This paradigm differed across the three experiments only in object-location pairings and precise memory tests. I) Encoding (in iVR): participants explored a virtual kitchen (45 sec) and were instructed to count and memorise 20 object locations. II) Recall (in iVR): all 20 objects were removed, and participants were given one object at a time to be placed where previously encoded. III) 3AFC (outside iVR): each of the 20 objects was presented on a computer screen but in three alternative locations (where the two foil locations were approximately equally expected as the correct location). Experiments 2 and 3 collected additional remember/familiar judgments. IV) Individual expectancy ratings (outside iVR), used in analysis of previous Recall and 3AFC data, were collected for all 20 objects for a) each of the 3AFC locations and b) their general expectancy in a kitchen context.

Recall was followed by a 3AFC recognition test, performed on a computer outside iVR. Each trial showed one studied object in three locations, one of which was correct. Importantly, the two foil locations were chosen to be approximately equally expected according to the normative ratings (see above), so using prior knowledge to guess the location would not help performance. Participants indicated which they thought was the studied location, followed by a rating of their confidence on a 3-point scale (1 = "*did not see the object*", 2 = "*guess the object was there*", 3 = "*know the object was there*"). In the final phase, participants provided *expectancy ratings* for how likely they thought it would be to find each of the 20 objects in each of the three locations tested in 3AFC, together with an additional rating of the *general expectancy* of an object appearing anywhere in a kitchen at all. Ratings were made using a sliding scale from unexpected (-100) to expected (+100). These ratings were analogous to the normative ratings, but allowed for potential individual differences in expectancy. The range of expectancy ratings for each participant and object (by number) is apparent in Figure 2.4.

Statistical analysis

Statistical analysis was performed in R (R Core Team, 2019) using Bayesian multi-level models with *brms* (Bürkner, 2017, version 2.12.0 2018) based on Stan (Carpenter et al., 2017). All analyses scripts and data are available here: https://osf.io/4sw2t/ and https://github.com/JAQuent/schemaVR.

Memory for individual trials was modelled as a function of a participant's object-location expectancy rating. Memory was a binary outcome (correct/incorrect) and was fitted using logistic regression models with the Bernoulli linking function. A "full" model was fit first, with random slopes and intercepts for both objects and participants. Bayes Factors (BFs) using marginal likelihoods from bridge sampling (Gronau et al., 2017) were then used to compare the full model with the model with random intercepts only, which was in turn compared to the model without random intercepts.

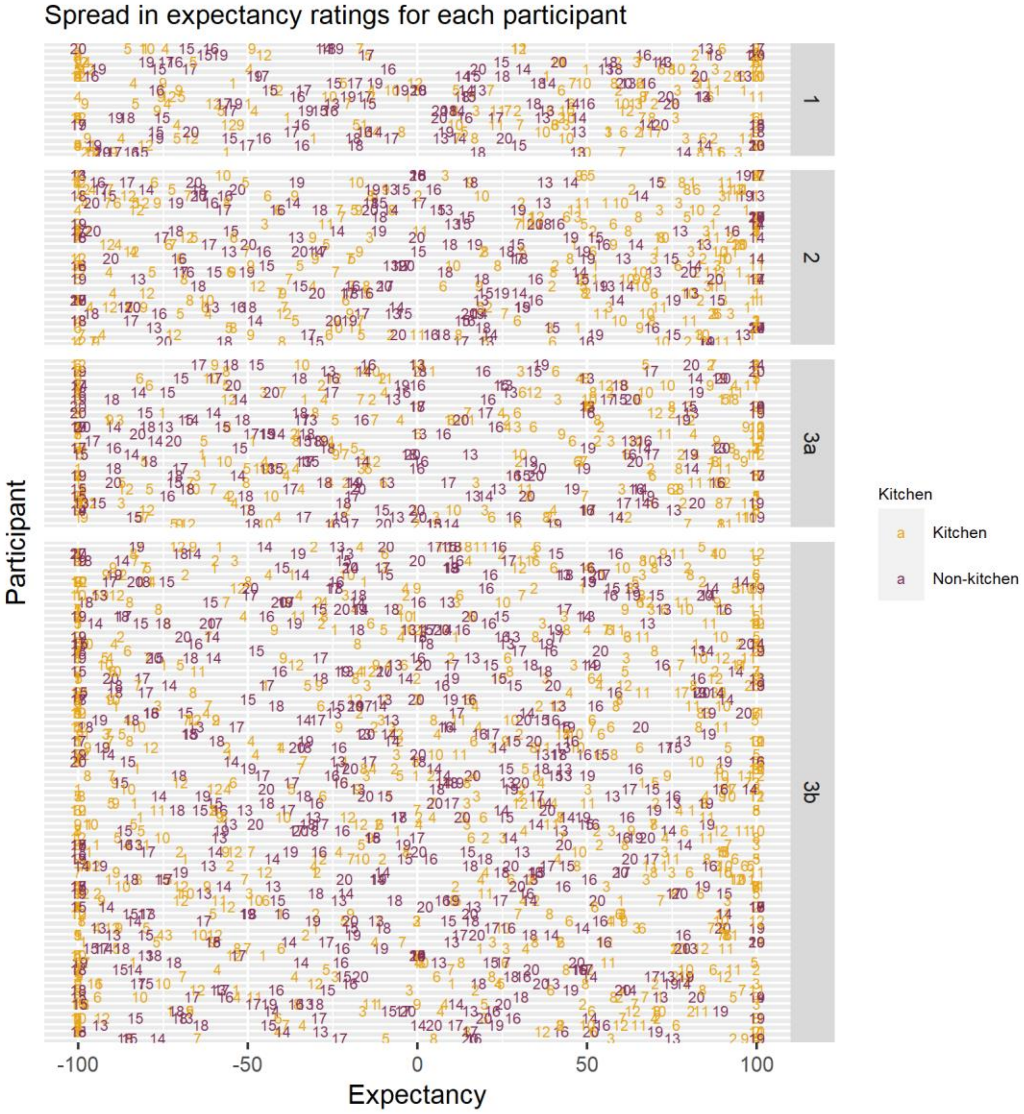

Figure 2.4: Illustration the spread of all object-location expectancy ratings: On the x-axis are the ratings collected from each participant (y-axis) tested across all three experiments (i.e. panel labelled 1, 2 and 3). Numbers represent the objects (see Table 1) and are coloured orange if they are generally expected in the kitchen and purple if they are not.

The individually-defined expectancy ratings (Figure 2.4) were scaled to have a standard deviation (SD) of 0.5, and the prior for each regression coefficient was based on a "Student t" distribution, with hyperparameters of df = 7, μ = 0, σ = 1, except for the intercept, which had hyperparameters of df = 7, μ = 0, σ = 10 (see https://jaquent.github.io/post/the-priors-that-i-use-for-logsitic-regression-now/ for justification; see also Table 7.1 ahead). These generic weakly informative priors are chosen to regularise unexpectedly large effects (Gelman et al., 2008). Eight Markov chain Monte Carlo (MCMC) chains were run, with 2000 warm-up and 16000 regular iterations and a total of 112,000 post-warmup samples for each main model. All models converged with an $\hat{R}$ of 1.

Evidence for or against my hypotheses were quantified with BFs for the linear and quadratic component of a second-order polynomial expansion of expectancy. A symmetrical U-shape would have a positive quadratic coefficient and a zero linear coefficient (see Experiment 3b for a more stringent test based on opposite signs of interrupted linear regression). The BF for each coefficient was estimated by the Savage-Dickey ratio (Wagenmakers et al., 2010). The test for the quadratic term was order-restricted (one-tailed) in line with my pre-registered hypothesis, all other tests were not order-restricted unless otherwise specified. For restricted tests, I compared the density of the truncated and renormalised prior distributions at zero with the logspline non-parametric density estimate of the truncated and renormalised posterior distributions of my parameters at zero (based on the 112,000 post-warm-up samples). For unrestricted tests, BFs were just density ratios at zero: of prior/posterior (BF_{10}) or posterior/prior (BF_{01}) without truncation and renormalisation. The Savage-Dickey ratio function used can be found in this GitHub repository: https://github.com/JAQuent/assortedRFunction. In addition to BFs, I report 95 % credible intervals around my parameters, which can be interpreted as evidence against the null hypotheses if they do not include zero.

Finally, when testing the means across trials, BFs were derived from Bayesian t-tests (Rouder et al., 2009) with the package 'Bayes factor' (version 0.9.12-4.2), with the default scale parameter of √2/2.

Results

Objects that were reported as 'not seen' were excluded from further analysis. Across both the recall and 3AFC task, a mean of 2.84 (SD = 1.96) objects were excluded.

For the recall data (Figure 2.5A), a model with random intercepts for participants and objects was used, since this was favoured relative to one that also included random slopes (BF = 1020) and relative to one that did not have random intercepts (BF = 1590). There was inconclusive evidence for a linear effect, β = 0.096 (95 % CI [-0.663, 0.885]), BF_{10} = 0.39, but more importantly, there was strong evidence for the predicted positive quadratic effect, β = 1.888 (95 % CI [0.315, 3.641]), BF_{10} = 25.04.

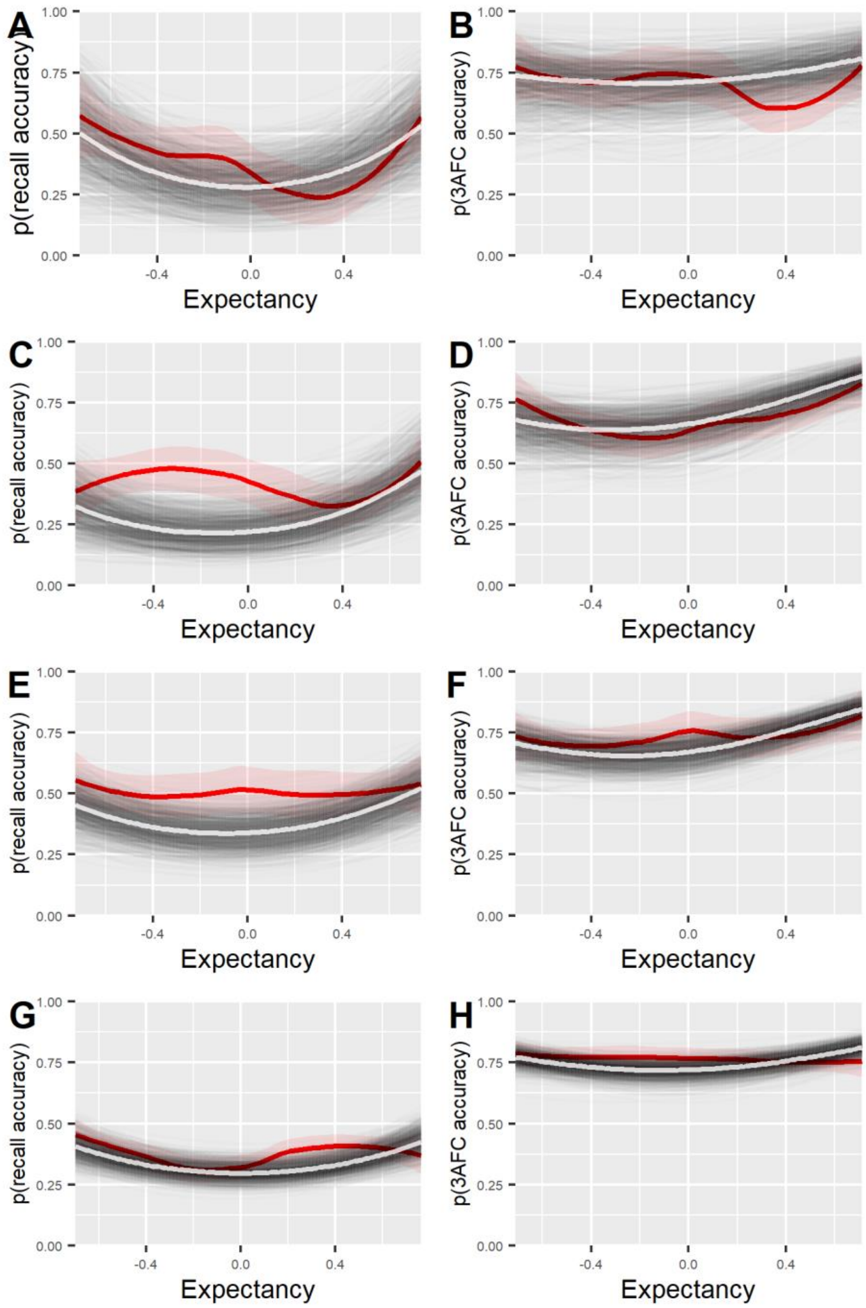

Figure 2.5: Accuracy results: Recall (left panels) and recognition (right panels) performance plotted across expectancy ratings for all four Experiments 1, 2, 3a and 3b (rows). The red line shows locally weighted smoothing (loess) of data to illustrate how well the fitted 2nd-order polynomial model using propagated evidence (white line) represents the data. The thin grey lines show 1,000 randomly selected fits from the posterior distributions, illustrating the uncertainty of the fit. Expectancy ratings originally ranged from -100 to +100, but were scaled to have SD = 0.5 in order to enable standard priors.

It is possible that the two sides of the U-shape arise at different stages, e.g., an advantage in encoding unexpected locations and an advantage in retrieving expected locations. To test the latter – i.e., whether expectancy influenced retrieval (e.g. guessing) – I compared the mean expectancy of incorrectly versus correctly recalled locations. Confirming expectations, the mean expectancy of incorrectly recalled locations was +40.7 (14.38), which was greater than zero, and greater than that for correctly recalled locations of +2.9 (10.6), BF_{10} = 23900, d = 2.19, supporting a bias for participants to report expected locations when unsure.

The 3AFC foils were designed to control for this potential bias towards recalling expected locations. Like for the recall data above, model comparison showed that the 3AFC data were better fit by a model with random intercepts but not random slopes, relative to the full model with random slopes (BF = 717) or the minimal model with no random intercepts or slopes (BF = 19.45). This model again showed inconclusive evidence for or against a linear effect, β = 0.29 (95 % CI [-0.48, 1.12]), BF_{10} = 0.49. However, in this case, the evidence for a quadratic effect was also inconclusive, β = 0.64 (95 % CI [-0.72, 2.07]), BF_{10} = 1.61 (Figure 2.5B).

In addition to 3AFC accuracy, high confidence (predicting 3 = "*know the object was there*" vs the other response options) was modelled with the same structure used above. Here, anecdotal evidence against a linear component, β = -0.008 (95 % CI [-0.651, 0.692]). BF_{10} = 0.331, and inconclusive evidence for a quadratic term, β = 1.009 (95 % CI [-0.25, 2.378]). BF_{10} = 2.01, was found.

Discussion

Experiment 1 confirmed the predicted U-shape function for recall of object locations as a function of the expectancy of those locations. However, this U-shape was not replicated when using accuracy from a 3AFC recognition test in which the foils were matched for expectancy, suggesting that some aspects of the U-shape might arise during retrieval rather than encoding, such as the bias towards guessing expected locations that was also observed. Yet, when predicting high confidence responses in the 3AFC there was at least inconclusive evidence in the right direction that there might be a U-shape. Nonetheless, the evidence from the 3AFC data was moot, and more data might be required to provide more compelling evidence.

Furthermore, a potential confound in Experiment 1 was that kitchen objects tended to be in highly congruent/incongruent locations, while non-kitchen objects primarily occupied by neutral locations. Even though random intercepts were allowed for each object, this meant that the U-shape might in part be explained by some theoretical difference between kitchen vs non-kitchen objects, other than location expectancy and other individual differences between objects. I addressed this in the next experiment.

2.4 Experiment 2

Experiment 2 was powered to have a better chance of detecting the predicted U-shape in 3AFC accuracy. The procedure was identical to Experiment 1, except that the specific object-locations studied were re-selected (from the normative ratings) to give a more even coverage across the range of expectancy values for both kitchen and non-kitchen objects.

The only other important difference is that participants were also asked to indicate the quality of their memory responses by using 'remember/familiar/guess' judgements (Gardiner et al., 2002; Rajaram, 1993). This was to test the second hypothesis of SLIMM, that the two ends of the U-shape would be associated with different types of memory, specifically recollection for unexpected locations and familiarity for expected locations. While it is traditionally assumed that recollection tends to be associated with high confidence and familiarity is not synonymous with low confidence (in that people can be very confident because of a strong feeling of familiarity in the absence of recollection) based on the seminal findings of Gardiner & Java (1990). However it can be questioned if that still holds given recent carefully pre-registered *failed* replication attempts (Haaf et al., 2021). Despite these concerns I decided that estimating recollection and familiarity are worthwhile in this case. Beyond that a final caveat concerns the fact that typically recollection and familiarity are measured for item recognition (Yonelinas, 2002) but here it is done for an associative memory type (object-location). This experiment was pre-registered on OSF (https://osf.io/s9er3).

Methods

Participants

I collected data from 25 new participants. These had a mean age of 24.52 (SD = 2.83) years, with 18 females, 6 males and 1 non-binary. Sample size was determined based on frequentist power analysis to achieve 80 % power (https://osf.io/gr98d/) based on the quadratic effect size (β = 0.51 on unit scale) for 3AFC (https://osf.io/s9er3).

Materials

The only change from Experiment 1 was that objects were now reshuffled to other locations, so that kitchen and non-kitchen objects were more evenly distributed across (normative) expectancy ratings. To address the problem that 'kitchen' objects tended to be in extreme positions, only the "SS of targets" for kitchen objects was used; otherwise the algorithm remained unchanged. This ensured that kitchen objects now also occupied middle locations, while the spread of non-kitchen objects was still adequate. I will call this Set 2, as distinct from the Set 1 used in Experiment 1.

Procedure

The only procedural change was how participants categorised their memory responses for both recall and 3AFC (replacing the previous categories based on confidence only): If they did not remember seeing the object at all, they were told to indicate "no memory". If they remembered the object itself, but had little idea where it was, they were to indicate "guess". If they did not initially remember where it was, but once they had placed it (in recall) or compared the three choices (in 3AFC), that location just looked familiar, they were to indicate "familiar". Finally, if they immediately remembered where the object was when they saw it (because, for instance, they remembered what they thought when they saw the object), then they were to indicate "remember".

Statistical analysis

Parts of the statistical analysis were identical to Experiment 1, with BFs based on the same zero-centred priors. However, I also used the posterior distributions of Experiment 1 as prior distributions by estimating the family-specific parameters of the Student's t-distribution for those distributions (see Table 7.1) using *brms*. For that, I used the marginal distribution, ignoring any correlation between parameters, and used the same factor to scale expectancy ratings as in Experiment 1, to ensure comparable expectancy ratings across experiments and hence correct priors (so the SD of these ratings in Experiment 2 was close to, but not identical to, 0.5). The BF from this second model allowed me to update the posterior belief in favour of my hypotheses (what I call "PB_{10}"), given the data from both experiments (calculated by multiplying this BF with the BF from the previous experiment).

Given the results of Experiment 1, random intercepts were modelled but not random slopes. Remember/familiar judgements were initially analysed in line with pre-registered analysis of the mean expectancy rating for remember and familiar judgments, but further simulation showed that this trial-averaged analysis is biased by boundary effects on expectancy values (see Appendix). Therefore, I analysed them using the same single-trial logistic regression model that I used for overall accuracy. To estimate the probability of recollection, an outcome of 1 was used for “remember” judgments, and an outcome of 0 otherwise. There is debate over the best way to estimate familiarity, i.e., whether familiarity and recollection are redundant, independent or exclusive (Knowlton & Squire, 1995). To model redundancy, familiarity was estimated with an outcome of 1 for “remember” or “familiar” responses, and 0 otherwise; to model independence, an outcome of 1 was used for “familiar” responses, but only trials that were not given a “remember” response were included; to model exclusivity, familiarity was estimated as for independence, and recollection was estimated by an outcome of 1 for “remember” responses, but only trials that were not given a “familiar” response were included.

Results

An average of 2.68 (SD = 2.52) objects were reported as ‘not seen’ and therefore excluded from further analysis. The distribution of individual expectancy ratings for kitchen and non-kitchen objects was also more evenly spread across expectancy (Figure 2.4).

For recall (Figure 2.5C), there was inconclusive evidence for a linear term, $BF_{10} = 3.25$, $\beta = 0.407$ (95 % CI [-0.074, 0.897]), $PB_{10} = 0.91$, and more importantly, continued very strong evidence for a positive quadratic term, $BF_{10} = 2.5$, $\beta = 1.538$ (95 % CI [0.445, 2.635]), $PB_{10} = 65.35$. Like Experiment 1, the average expectancy for incorrectly-placed objects was +34.49 (16.59) and so clearly higher than for correctly-placed, which was -4.54 (14.58), $BF_{10} = 174000$, $d = 1.51$

Focusing on the 3AFC therefore, there was a clear linear term, $BF_{10} = 104.69$, $\beta = 0.749$ (95 % CI [0.26, 1.252]), $PB_{10} = 22.73$, and more importantly, a clear quadratic term, $BF_{10} = 6.71$, $\beta = 1.169$ (95 % CI [0.18, 2.182]), $PB_{10} = 13.72$. The positive linear term produced an asymmetry in the U-shape towards expected locations (Figure 2.5D).

Assuming that recollection and familiarity are independent, the estimate of recollection from 3AFC responses (Figure 2.6A) showed inconclusive evidence for a linear term, $BF_{10} = 0.46$, $\beta = 0.292$ (95 % CI [-0.212, 0.81]), but strong evidence for a quadratic term, $BF_{10} = 19.29$, $\beta = 1.479$ (95 % CI [0.394, 2.633]). By comparison, the estimate of familiarity (Figure 2.6B) showed anecdotal evidence against a linear term, $BF_{10} = 0.31$, $\beta = 0.099$ (95 % CI [-0.497, 0.705]), but inconclusive evidence for and against a quadratic term, $BF_{10} = 0.60$, $\beta = 0.102$ (95 % CI [-1.119, 1.35]). For the recollection estimate under exclusivity, there was no linear effect, $BF_{10} = 0.50$, $\beta = 0.322$ (95 % CI [-0.3, 0.971]), and anecdotal evidence for a quadratic effect, $BF_{10} = 4.17$, $\beta = 1.317$ (95 % CI [0.005, 2.756]).

For familiarity scored under the redundancy assumption, there was anecdotal evidence against a linear effect, $BF_{10} = 0.33$, $\beta = 0.186$ (95 % CI [-0.35, 0.743]), but inconclusive evidence for a quadratic effect, $BF_{10} = 2.18$, $\beta = 0.962$ (95 % CI [-0.145, 2.141]).

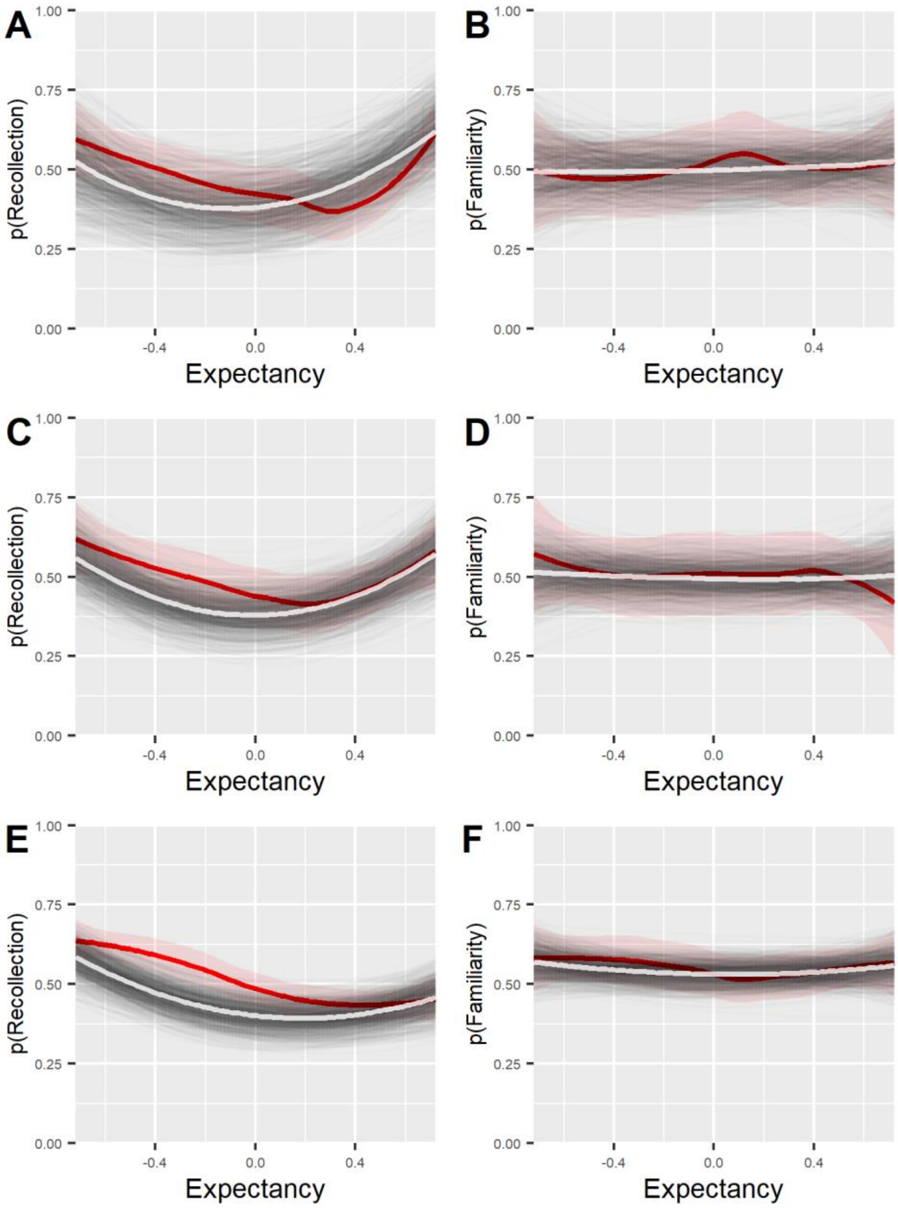

Figure 2.6: Recollection & familiarity results: 3AFC recognition performance based on recollection (under independent or redundant scoring, left panel) and familiarity (under independent or exclusive scoring, right panel) estimates across Experiment 2, Experiment 3a and Experiment 3b (rows). The red line shows locally weighted smoothing (loess) of data to illustrate how well the average model (the white line) represents the data. The thin grey lines show 1,000 randomly selected fits from the posterior distributions, to illustrate the uncertainty of the average model.

Expectancy originally ranged from -100 to +100, but were scaled to have a SD = 0.5 in order to enable standard priors.

Discussion

Experiment 2 replicated the U-shape in recall of Experiment 1, using a more even distribution of kitchen/non-kitchen objects across expectancy values, but now also found a U-shape in 3AFC, which rules out a contribution from guessing expected locations. This confirmed the first of my original hypotheses. However, the results provided strong evidence against my second hypothesis, that the advantage for unexpected locations would be accompanied by recollection, while that for expected locations would be accompanied by familiarity. Rather, I found that both ends of the U-shape were accompanied by recollection, whereas familiarity showed little variation as a function of expectancy, regardless how recollection and familiarity were scored.

One limitation of Experiments 1-2 is that the same set of 20 object-location pairings were used for all participants (even if they were changed slightly across experiments), raising the possibility that the U-shape was a quirk of these particular pairings. Therefore, for Experiment 3a and 3b, I created five new sets of object-location pairings, counterbalanced across participants, to check whether the U-shape generalised.

2.5 Experiment 3a

Experiment 3a was run to check whether I could replicate the U-shape again, particularly in 3AFC, but this time across a large range of object-location pairings. For this, five distinct sets (i.e. unique object-location pairing pairings) were created and to which participants were randomly assigned.

Methods

Participants

25 participants were recruited to take part in the study, but one was excluded from further processing due to experimenter error by which they rated different stimuli to those presented in the study. Hence a total of 24 participants (11 females and 13 males, mean age 25 years, SD = 3.71 years) were included in the analysis. This means that the data is

not completely counterbalanced across the 5 stimulus sets. This experiment was pre-registered at (https://osf.io/kcr2q).

Materials, Procedure & Statistic analysis

Methods were identical to Experiment 2, except that five new sets of 20 object-location pairings were based on the normative data (Sets 3-7; see Appendix Table 7.2), each chosen to maximise the range of expectancy values (materials available here https://osf.io/4sw2t/ and https://github.com/JAQuent/schemaVR).

Results

An average of 2.12 (SD = 1.53) objects were reported as 'not seen' and excluded from further analysis.

For recall (Figure 2.5E), the results showed anecdotal evidence against a linear term, BF_{10} = 0.22, β = 0.189 (95 % CI [-0.144, 0.522]), PB_{10} = 0.30, but very strong evidence for a quadratic term, BF_{10} = 1.33, β = 1.15 (95 % CI [0.351, 1.945]), PB_{10} = 59.47. Like Experiment 1 and 2, the average expectancy for incorrectly-placed objects M = +37.39 (SD = 15.69) was clearly higher than for correctly-placed objects M = +1.08 (SD = 14.18), BF_{10} = 663000, d = 1.7.

For 3AFC (Figure 2.5F), there was very strong evidence for a linear term, BF_{10} = 0.88, β = 0.59 (95 % CI [0.233, 0.949]), PB_{10} = 41.82, but there also strong evidence for a quadratic term, BF_{10} = 2.32, β = 1.137 (95 % CI [0.348, 1.928]), PB_{10} = 43.09.

For recollection estimates from 3AFC under independence scoring (Figure 2.6C), there was moderate evidence against a linear term, BF_{10} = 0.26, β = 0.026 (95 % CI [-0.311, 0.36]), PB_{10} = 0.166, but more importantly, extreme evidence for a quadratic term, BF_{10} = 5.74, β = 1.475 (95 % CI [0.682, 2.27]), PB_{10} = 506.17. For familiarity (Figure 2.6D), there was anecdotal evidence against a linear, BF_{10} = 0.36, β = -0.177 (95 % CI [-0.793, 0.445]), PB_{10} = 0.22, as well as inconclusive evidence either way regarding the quadratic term, BF_{10} = 0.61, β = 0.092 (95 % CI [-1.127, 1.335]), PB_{10} = 0.48.

For recollection estimated under exclusivity, I found no linear effect, BF_{10} = 0.26, β = 0.027 (95 % CI [-0.393, 0.447]), PB_{10} = 0.205, and very strong evidence for a quadratic effect, BF_{10} = 5.74, β = 1.442 (95 % CI [0.468, 2.428]), PB_{10} = 59.1. For familiarity scored under the redundancy assumption, there was no linear effect, BF_{10} = 0.34, β = -0.047 (95 % CI [-0.422, 0.327]), PB_{10} = 0.185, but moderate evidence for quadratic effect,

$BF_{10} = 1.55$, $\beta = 1.009$ (95 % CI [0.169, 1.854]), $PB_{10} = 6.69$, which can be traced back to the contribution of recollection.

Discussion

In general, Experiment 3a confirmed the U-shape function for both recall and recognition as a function of object-location expectancy, particularly when combined with Experiments 1 and 2, and confirmed that both ends of this U-shape were associated with recollection. However, the BF for a positive U-shape in 3AFC (regardless of recollection/familiarity) was not strong, $BF_{10} = 2.32$, which may reflect weaker effects for some of the new stimulus sets, and there was also continued evidence (when combined across all experiments) for an accompanying positive linear effect for 3AFC, which could swamp a true U-shape (see Figure 2.5F). Therefore I decided to collect more data with these stimulus sets, so as to allow a more stringent test of a U-shape: namely, interrupted regression that tests whether both ends of the U-shape are independently reliable (see later). These data were collected as part of the novelty experiment in Chapter 3.

2.6 Experiment 3b

Experiment 3b was identical to Experiment 3a, except that the participants had previously studied a list of unrelated words for the novelty experiment described in Chapter 3.

Methods

Participants

A total of 72 participants (50 females, 21 male and 1 non-binary, mean age 26.12 years, SD = 6.53 years) were tested, counterbalanced across the 5 stimulus sets. Note that the sets were not fully counterbalanced across participants (see later). Prior to this experiment, the participants had previously also made judgments about a list of 288 words, as part of a separate Registered Report examining the effect of the subsequent iVR experience on incidental memory for those words (see Chapter 3). The words were not related to kitchens or the objects used in the iVR phase, and participants were told the words were not relevant to the iVR phase. This experiment was also pre-registered at OSF (https://osf.io/b9dqg).

Results

An average of 2.31 (SD = 1.65) objects were reported as 'not seen' and excluded from further analysis.

For recall (Figure 2.5G), the results showed inconclusive evidence against a linear term, BF_{10} = 0.21, β = -0.007 (95 % CI [-0.225, 0.208]), PB_{10} = 0.11, but extreme evidence for a quadratic term, BF_{10} = 8.27, β = 0.991 (95 % CI [0.462, 1.522]), PB_{10} = 553.68. Like Experiment 1, 2 and 3a, the average expectancy for incorrectly-placed objects M = +23.44 (SD = 22.33) was clearly higher than for correctly-placed objects M = -9.06 (SD = 25.32), BF_{10} = 1.46e+10, d = 1.09.

For 3AFC (Figure 2.5H), there was inconclusive evidence for or against a linear term, BF_{10} = 0.24, β = 0.177 (95 % CI [-0.067, 0.417]), PB_{10} = 0.42, but strong evidence remained for a quadratic term, BF_{10} = 0.8, β = 0.8 (95 % CI [0.227, 1.37]), PB_{10} = 25.85.

For recollection estimates from 3AFC under independent scoring (Figure 2.6E), there was extreme (additional) evidence for a linear term, BF_{10} = 1545.95, β = -0.353 (95 % CI [-0.568, -0.144]), which held even when combined with previous experiments, PB_{10} = 23.12, but also, strong evidence for a quadratic term, BF_{10} = 0.8, β = 0.949 (95 % CI [0.425, 1.473]), PB_{10} = 258.15. For familiarity under independent scoring (Figure 2.6F), there was anecdotal evidence against a linear, BF_{10} = 0.19, β = -0.037 (95 % CI [-0.324, 0.254]), PB_{10} = 0.20, an inconclusive evidence regarding the quadratic, BF_{10} = 0.51, β = 0.245 (95 % CI [-0.414, 0.906]), PB_{10} = 0.42, term. For recollection estimate under exclusivity, I found a strong linear effect, BF_{10} = 1550, β = -0.361 (95 % CI [-0.641, -0.085]), PB_{10} = 8.84, and moderate evidence for a quadratic effect, BF_{10} = 0.8, β = 1.097 (95 % CI [0.43, 1.762]), PB_{10} = 7.55. For familiarity scored under the redundancy assumption, there was inconclusive evidence regarding a linear effect, BF_{10} = 4.28, β = -0.25 (95 % CI [-0.497, -0.005]), PB_{10} = 1.52, but moderate evidence for quadratic effect, BF_{10} = 0.9, β = 0.792 (95 % CI [0.214, 1.372]), PB_{10} = 3.68.

The combined evidence across experiments (see PB) for a quadratic term in the absence of a linear effect is consistent with a U-shaped function. However, a stronger test is to check that both ends of a U-shape are reliable, by testing for opposite linear slopes; so-called "interrupted regression" (e.g. Simonsohn, 2018). As pre-registered (see https://osf.io/b9dqg), I tested this by combining data across all experiments. To determine slopes of interrupted linear regressions, one must find a "breaking-point" (bp) somewhere along the continuum, at which the two slopes meet. This was done by fitting these two models:

y ~ xlow + xhigh + high + set[3] + (1 | participant number) + (1 | object number)

where:

For Slope 1:

x = object/location expectancy that is transformed to

xlow = x – bp if x <= bp and 0 otherwise,

xhigh = x – bp if x > bp and 0 otherwise, and

high = 1 if x > bp and 0 otherwise.

For Slope 2:

x = object/location expectancy that is transformed to

xlow = x – bp if x < bp and 0 otherwise,

xhigh = x – bp if x >= bp and 0 otherwise, and

high = 1 if x >= bp and 0 otherwise.

These models were applied to 10 equally-spaced breaking points within the middle 80% range of the expectancy ratings. Simulations showed that the false positive rate remained under 5% for testing these 10 breaking points for accepting a $BF_{10} > 6$ as evidence (see https://jaquent.github.io/post/finding-a-u-shape-with-bayesian-interrupted-regression/).

For recall, the strongest effect was with a breakpoint of +0.06 (9.24 in original scale), where the estimated leftward slope (towards negative expectancy values) was β = -0.78

[3] Note that I originally pre-registered to model experiment as well as set but this was mistake as Set 1 and Set 2 would be modelled twice. Now, set is modelled as a combination of set and experiment leading to a regressor that has 12 levels (Set 1, Set 2, Set 3 for Experiment 3a, …, Set 3 for Experiment 3b, …).

(95 % CI [-1.30, -0.28]), BF_{10} = 54.89, and the rightward slope was β = +0.69 (95 % CI [0.06, 1.34]), BF_{10} = 6.27. For 3AFC, the best breakpoint of -0.26 (-34.3 in original scale) had a leftward slope of β = -0.84 (95 % CI [-1.77, 0.12]), BF_{10} = 3.93, and a rightward slope of β = +0.50 (95 % CI [0.10, 0.92]), BF_{10} = 8.48. Note though that the leftward slope fell below the BF_{10} criterion that I set (BF_{10} > 6). For recollection estimates of 3AFC, the best breakpoint of +0.22 (30 in original scale) had a leftward slope of β = -0.84 (95 % CI [-1.23, -0.45]), BF_{10} = 619.55, and rightward slope of β = +1.22 (95 % CI [0.37, 2.10]), BF_{10} = 40.45. These results provide strong support for both sides of the U-shape being reliable in recall and for recollection in particular. Full results, including for familiarity that did not show any indication of U-shape, can be found in Table 7.3 to Table 7.6 in the Appendix.

Effect of Stimulus Set

Even though PB remained supportive for a U-shape in the critical 3AFC test, the BFs for Experiment 3a and 3b separately were weaker than in Experiment 2 (and to a lesser extent, than in Experiment 1). This raises the possibility that some sets of object-location pairings give stronger U-shapes than others.

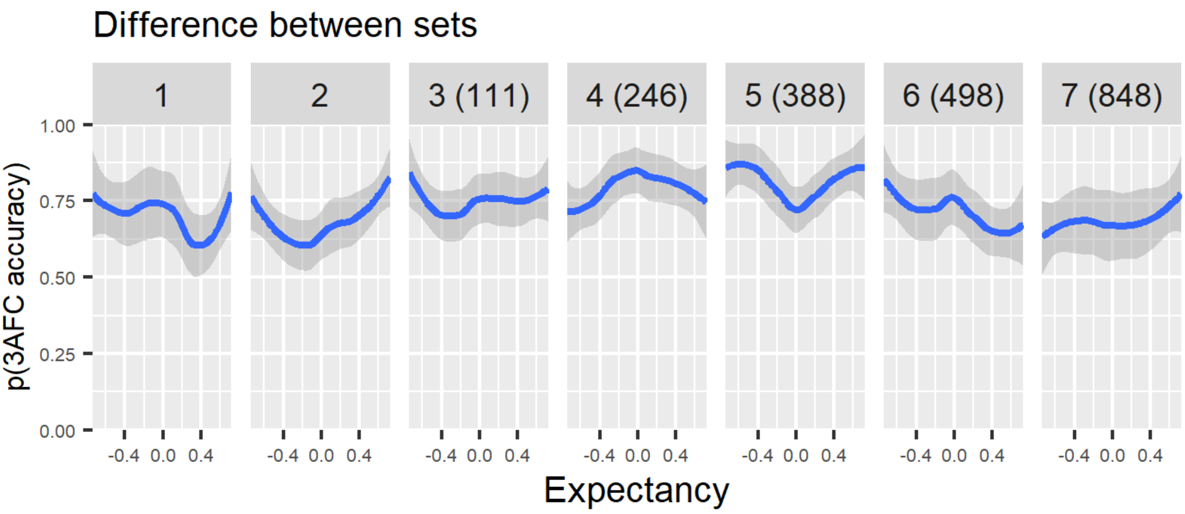

Figure 2.7: 3AFC accuracy as a function of set: Blue line represents the LOESS line to illustrate the ‘raw’ relationship between expectancy and 3AFC accuracy. The number in parenthesis represents the set number in the data.

To test this, I combined data from all Experiments, treating Experiment 1 and Experiment 2 as their own sets. Figure 2.7 suggests that there was variability between the sets with respect to whether they produce a U-shape. First I examined whether the distribution of expectancy values varied across the different sets. For this, I binned the expectancy values into 8 bins and calculated the frequency of assignment for each bin. Then, I calculated the absolute difference for each bin from the average of all bins. This analysis revealed that set 4 (246), the set that seems to show an inverted U-shape, and set 5 (388), the set that eyeballing showed the strongest U-shape, have very similar values, suggesting that the stimulus sets do not differ much in their expectancy distribution. Second, I inspected each set with to check whether any particular idiosyncratic pairing of object/location (e.g. an unexpected or expected pairing that is difficult to see and therefore associated with low accuracy) might explain the differences; however I could find no obvious explanation.

Finally, in order to formally investigate this question, I fitted three models using the pooled data from all experiments. The first model also allowed the factor "set" with seven levels as to interact with the linear and the quadratic component of expectancy; the second model included the factor set only as a fixed effect; the third model did not include the set factor at all. The function bayes_factor() was used to compare the different models, which all used zero-centred priors as previously described. These models were run with all measures (recall, 3AFC and recollection) that exhibited a U-shape in the previous iterations of the experiment. For those, the model including set was favoured over the model that also included interactions between set and expectancy (recall BF_{10} = 2290, 3AFC BF_{10} = 15.2, recollection BF_{10} = 109). At the same time, the simplest model that did not include set at all was better still, for all measures (recall BF_{10} = 129, 3AFC BF_{10} = 13.9, recollection BF_{10} = 229). Because there was no evidence that the stimulus set affected performance, particularly in interacting with expectancy (despite the numerical differences in Figure 2.6), I think I am justified in propagating evidence across experiments, as done in this chapter.

Pooled vs Propagated Evidence

To check that the propagation of Bayesian evidence was itself accurate, despite ignoring any posterior dependencies between parameter values, I checked that the BF from the pooled data was similar to the final PB after Experiment 3b. For recall, the pooled estimate of the linear effect (β = 0.07, 95 % CI [-0.13, 0.27]) had a BF_{10} = 0.12 which is comparable to PB_{10} = 0.35 reported for Experiment 3b above; while the pooled estimate of the quadratic effect (β = 0.957, 95 % CI [0.464, 1.452]) had a BF_{10} = 306, again comparable to the above PB_{10} = 554. For 3AFC, the pooled estimate of the linear effect (β = 0.182, 95 % CI [-0.035, 0.4]) had a BF_{10} = 0.415, comparable to the above PB_{10} = 0.42, while the pooled estimate of the quadratic effect (β = 0.798, 95 % CI [0.278, 1.32]) had a BF_{10} = 23.2, comparable to the above PB_{10} = 25.85. Finally, for recollection, the pooled estimate of the linear effect β = -0.33, 95 % CI [-0.529, -0.132]) had a BF_{10} = 18.4, comparable to the above PB_{10} = 23.1, while the pooled estimate of the quadratic effect (β = 0.95, 95 % CI [0.465, 1.439]) had a BF_{10} = 370, comparable to the above PB_{10} = 258. Finally, I also confirmed that reversing the order of calculating PB, i.e. using posteriors from Experiment 3b as priors on Experiment 3a, etc, produced exactly the same final PB value, as expected.

Discussion

Experiment 3b generally confirmed the U-shape function for both recall and recognition as a function of object-location expectancy, particularly when combining evidence with Experiments 1, 2 and 3a, and confirmed that both ends of this U-shape were associated with recollection. This was apparent by strong evidence for a positive quadratic component, and in more strict interrupted regression analyses, which provided anecdotal to strong evidence that both sides of the U-shape showed a significant linear effect of opposite sign. The weakest evidence was a BF_{10} of 3.93 for the leftward slope for 3AFC, i.e., the unexpected side of the U-shape, which fell just below the criterion of at least BF = 5.66 that simulations[4] showed controlled the false positive rate for the 10 break points tested. Nonetheless, the finding that there was evidence against a linear component to the overall fit (only evidence for a quadratic component) suggests that the dependency of 3AFC on expectancy really is U-shaped. Interestingly, the overall fit for recollection did show evidence for a negative linear effect (in addition to a positive quadratic component), suggesting that, while the interrupted regression confirmed that both ends of the recollection dependency were reliable, there was an additional bias towards better recollection of unexpected locations.

2.7 General Discussion

I was able to confirm my first prediction that memory can be a U-shaped function of the expectancy of an occurrence, with better memory for highly expected or highly

[4] https://jaquent.github.io/post/finding-a-u-shape-with-bayesian-interrupted-regression/

unexpected information. This prediction was derived from the SLIMM model that presumes that different brain systems underlie the two ends of the expectancy dimension, and reconciles many previous studies that have claimed superior memory for either expected (schema-congruent) information or unexpected (surprising) events: whether one finds one advantage or the other depends on the relative position of the two experimental conditions along the expectancy continuum. Only by exploring a continuous and extreme range of expectancy can one see the larger picture.

The present finding replicates and extends the previous demonstration of a U-shape using three experimental conditions (Greve et al., 2019), and importantly demonstrates that the U-shape also occurs using rich, pre-experimental knowledge, rather than a simple rule learned during an experiment. Indeed, my use of iVR allowed me to test memory quickly (less than a minute of encoding) in a realistic situation, while simultaneously providing the precise experimental control needed to measure memory.

The present results are at odds with another study that featured three ordinal conditions that differed in expectancy (van der Linden et al., 2017). In that study, participants viewed sequences containing four object images, where all four belonged to the same theme (i.e. expected condition), or did not follow a particular theme (i.e. baseline condition), or the first three objects followed a theme but the fourth object violated it (i.e. unexpected condition). No differences between conditions were found in an immediate item recognition task, but after 24 hours, recognition performance for the fourth object in the expected condition was better than in the other two conditions. Interestingly, this result is unlikely to be due to schema-congruent guessing, because lures that were from the same theme as the schema-congruent targets were not falsely recognised more often than lures that were not related to any theme that was used.

This lack of any advantage for van der Linden et al.'s unexpected condition relative to their baseline condition is difficult to reconcile with my findings and with SLIMM. One possible reason could be that their unexpected condition was simply not potent enough. Another possibility relates to their recognition task, which might have been based more on familiarity than recollection (particularly perceptual familiarity, since perceptually-similar lures did show increased false alarms). The "local PE" of PIMMS predicts that PE increases associations between the predictor (preceding three objects) and the (un)predicted object, but not necessarily memory for the object itself (so would predict an advantage in unexpected condition relative to baseline if associative memory were tested instead, e.g., "what stimulus came after this one?"). Any transient changes between

PIMMS's item and feature levels (which underlie familiarity) would not necessarily differ across conditions. The "global PE" of SLIMM, on the other hand, would predict that the surprising object in the unexpected condition triggers an episodic snapshot, which should improve recollection of that object (by associating it with co-occurring context), but it is still possible that recognition was dominated by familiarity, so such improved recollection played little role. This "item recognition" contrasts with my 3AFC task, which explicitly tested the association between an object and its location in the kitchen (see Chapter 5 for further discussion).

My results may reconcile those from previous studies that have examined memory for objects within familiar types of rooms. In their seminal work, Brewer and Treyens (1981) reported that memory was positively correlated with schema expectancy. However, the opposite effect, where atypical objects in a room were remembered better, has also been reported (Lampinen et al., 2000, 2001; Pezdek et al., 1989; Prull, 2015). The discrepancy between these studies most likely reflects two factors: 1) with only restricted range of expectancy, the relative level of memory depends on the position of those points on the U-shaped continuum (as noted above), and 2) the precise type of memory assessed. Both of these issues apply, for example, to the study of Lew and Howe (2017), which is most similar to the present study, except that it used realistic photos of familiar room-types[5]. Like here, the photos showed room-congruent objects at expected and unexpected locations, as well as room-incongruent objects. At test, objects either stayed in same location or shifted to a different location. Recognition of objects was better for room-incongruent objects as well as room-congruent objects at unexpected locations, relative to room-congruent objects at expected locations. However, as noted above, their

5 The only previous study to investigate memory for objects using iVR, of which I am aware, was by Draschkow & Võ (2017). These authors examined how room schemas influence interactions with, and memory for, objects. Participants were asked to either arrange objects in a virtual room in a manner that was consistent or inconsistent with their expected arrangement. Participants spent more time handling and searching for inconsistently arranged objects, and consistently arranged objects were recalled more accurately than inconsistently arranged objects (possibly reflecting the guessing bias described above). While this study is a clever illustration of the use of iVR to study the effects of schema, it did not explicitly examine the continuous relationship between expectancy and memory.

recognition test was likely also influenced by memory for the object themselves, whether or not their location was switched (whereas here, I ignored the rare trials when an object was not recognised at all, and my "recognition" refers to recognition of the object's location). Indeed, when Lew and Howe (2017) tested memory by recall instead, memory for room-congruent objects at unexpected locations was impaired relative to room-congruent objects at expected locations. The authors explained this in terms of schemas acting differently on item and associative (location) memory, whereby an unexpected location attracts attention, but also activates schema-congruent bindings that interfere with memory (see also Bower et al., 1979). However, an alternative explanation for their dissociation between recognition and recall is that their recall test, but not recognition test, was influenced by guessing of room-congruent locations when memory failed (Bayen et al., 2000; Kleider et al., 2008; Konopka & Benjamin, 2009; Kuhlmann et al., 2012). This is indeed what I found when examining errors in my recall test. By controlling for this guessing bias in my 3AFC, through matching the expectancy of the foils, I was able to reveal an additional advantage of memory for unexpected locations. Thus, it appears critical to not only explore a range of expectancy values, but also control for other ways in which schema can influence memory performance, such as at retrieval as well as encoding.

With regard to the second prediction derived from SLIMM, namely that one side of the U-shape (high unexpectancy) would be associated with recollection of contextual details, while the other side (high expectancy) would be associated only with a feeling of familiarity (see also Kafkas & Montaldi, 2018), my evidence does not support this prediction. Indeed, my Bayesian analysis provided evidence against any substantial effect of expectancy for familiarity, and evidence instead for a U-shaped function of recollection. One possibility is that the SLIMM conception of the relationship between expectancy/schema-congruency and recollection versus familiarity is incorrect. Another possibility is that my assessment of memory for the location of objects (rather than the objects themselves; see above) inherently requires the same associative mechanisms that support recollection (and which are assumed to be supported by the hippocampus/MTL system). One way to test the latter would be to examine memory for the objects themselves (such as their perceptual details), rather than their location, and see if such "item memory" is associated with recollection when the object was highly unexpected, but familiarity when the object was highly expected. However, an additional possibility is that the U-shape arose because I tried to measure recollection and familiarity for a relatively unorthodox scenario (object-location memory) and that participants merely conflated recollection with high confidence responses.

My study raises interesting future questions, though also possible limitations. One interesting question is whether the U-shape applies to all types of expectancy, or only predictions deriving from pre-existing knowledge, i.e. schema. For example, would the same U-shape emerge for events that are expected or unexpected given an episodic context, such as a temporal sequence of items that enable a prediction for the next item (e.g. Kumaran & Maguire, 2006, 2007, 2009). In this context, it is important to differentiate the two dimensions of novelty (never seen – completely familiar) and of expectedness (unexpected – expected). These dimensions intersect in a non-orthogonal pattern, which implies that they are also not parallel or identical. For instance, given a certain context a familiar object can be unexpected and furthermore something can be rare but still experienced before hence generally familiar. Based on this distinction, I expect a U-shape only for the dimension of expectedness and even then merely if schematic knowledge is strong enough so it can help encoding. Based on this, I hypothesise one could observe a U-shape for a match vs. non-match paradigm with three conditions (order violation, random sequence and no order violation) provided that the sequences are studied well. This idea is lines well with the finding that the hippocampus is especially sensitive to mismatch (e.g. Kumaran & Maguire, 2006, 2007, 2009).

However, while others have shown a U-shape for verbal rules (Greve et al., 2019), the same may not apply to other stimuli and other types of memory test. A second limitation of the present study is that I only tested immediate memory, and it is possible that the U-shape changes as the retention interval increases, e.g., following consolidation processes that might occur overnight. Either way, my results demonstrate the important role of prior knowledge in shaping the encoding of new memories, and unify two factors (schema-congruency versus surprise) that have previously tended to be studied separately.

3 Effect of novelty

3.1 Introduction

In Chapter 1, I discussed how BTT describes the phenomenon that experiencing something novel can boost memory for surrounding information, and how this phenomenon is not explained by PIMMS. One of the BTT's claims, derived mainly from animal data, is that only initially weak memories benefit from this "novelty boost". Furthermore, another hypothesis is that, at least for spatial novelty, hippocampus-dependent memory should benefit the most, which might translate to enhanced recollection, though could in principle also lead to increased familiarity if we assume that familiarity typically represents weaker memory. In this chapter, I test hypotheses in humans.

To do this, I attempted to demonstrate a retroactive effect of a novel spatial navigation experience on memory for unrelated words, as a function of the encoding task (deep vs. shallow) and retrieval quality (recognition with remember vs. know judgments, plus recall). For that, one half of the words were encoded deeply using an animate/inanimate task (like in Fenker et al., 2008), while the other half were encoded shallowly using an orthographic task, which results in worse memory, i.e. weaker encoding (Craik & Lockhart, 1972; Otten et al., 2001; Yonelinas, 2002). Encoding was incidental, i.e. I did not tell participants that their memory for the words would be tested later. Like Schomaker et al (2014), I also used VR, but in particular an immersive VR system in which participants can physically walk around a virtual room (rather than navigating with a controller, as in Schomaker et al., 2014). Due to the current rarity of immersive VR systems (compared to video games or even passive VR), I expect this to be a highly novel experience (and I excluded people who have experienced immersive VR before). Indeed, in my prior work with immersive VR (see Chapter 2), most participants were amazed by their experience. Immersive VR also renders the novel experience more similar to the open-field, spatial exploration used to induce novelty in non-human animals. To isolate the novelty of the experience from the sensory, motor and executive demands of the VR task, the control group had experienced the same VR task the day before, so it was no longer novel. In summary, according to BTT, I expected to find: 1) a basic novelty effect (better memory for the preceding words in the novel group vs control group), 2) a greater novelty effect for shallowly than deeply encoded words, and 3) a novelty effect that is either larger or smaller for recollected words (Remember judgments) relative to words judged as familiar (Know judgments), depending on whether any boost is restricted to hippocampal-dependent memory, or whether any boost is restricted to weaker forms of memory.

3.2 Methods

Participants

All participants were recruited from MRC Cognition and Brain Sciences' SONA system, in-house participant panel or word-to-mouth. This study was approved by the Cambridge Psychology Research Ethics committee (PRE.2018.107). Data collection was stopped after data from 72 participants led to $BF_{10} > 6$ or $< 1/6$ for one of my planned comparisons (see Results), as registered. Participants were paid £6/h and received up to £3 for travel compensation per visit. Full payment was only made after successful completion of Day 3. Note these were the same participants used in Experiment 3b of Chapter 2.

Data collection had to be paused due to the COVID-19 pandemic. Approximately half of the participants completed the task before the pandemic and the rest after. The group ratio (novelty vs. control) at the beginning of the pandemic was circa 2.2 to 1 (and therefore reversed for participants tested after the COVID lock-down).

82 participants were tested in total. Ten were excluded for the following reasons: three participants did not complete the recognition task the next day, two had technical failures, two datasets were invalid due to experimenter error, one felt unwell after Day 1, one had prior VR experience (a registered exclusion criterion) and one participant's Pr was at

chance level, as defined by the bootstrapping procedure.[6] A further four participants were included, except for tests for which they had missing data: one had missing recall data and the other missed the VR questionnaire. Furthermore, two participants did not use the correct keys so their data had to be excluded from the analysis of the encoding task.

The final sample size of 72 participants contained 53 females, 18 males, 1 non-binary, with mean age M = 26.3 years (SD = 6.25 years). Participants in the Novelty group completed the online recognition task M = 25.5 hours (SD = 2.38 hours) after encoding, while the Control group completed the task after M = 25.1 hours (SD = 2.17 hours).

Procedure

In terms of which stimuli (see below) and the procedure I closely followed a previous paper by Otten et al. (2001), ensuring that memory performance was in the correct ballpark. The experiment ran over 3 days (see Figure 3.1), with Days 1-2 in the laboratory and Day 3 at home. On the critical Day 2, the experimental procedure can be divided into three phases: study, iVR and test phase. In the study phase, participants incidentally encoded words. The words were presented in four blocks, with one of the two study tasks (see below) in each block (counterbalanced across participants as ABBA or BAAB). Then they performed the iVR task (details below). The only difference between the Novelty and the Control group is that the Control group had already completed the iVR task the day before, but both completed a memory task in iVR. Finally, they freely recalled as many of the words as possible and completed a short questionnaire about the iVR experience. On the final Day 3, participants performed a recognition task with remember/know/new judgments to distinguish studied versus new (unstudied) words. I decided to test recall only on Day 2 and recognition memory only on Day 3 to minimise retrieval-induced enhancement or forgetting (M. C. Anderson et al., 1994).

[6] Note that I originally planned to remove participants with boot-strapped chance performance on the MPT parameters *r* and *f*. However some participants did not use both response options often enough to establish chance levels, so for the purpose of participant exclusion, I switched to the simpler Pr measure, collapsed across response options.

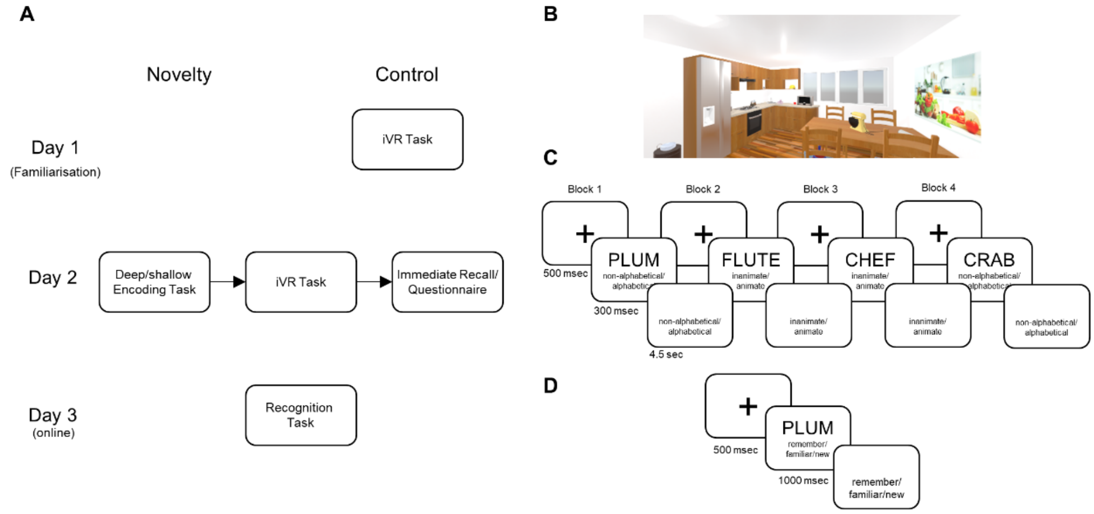

Figure 3.1: Task design: Panel A Illustration of experimental design: On day 1, the Control group was familiarised with the immersive virtual reality (iVR) task (see Panel B). Day 2 is the same for both the Novelty and the Control group. It started with the deep/shallow encoding task (see Panel C), immediately followed by the iVR task. This task was novel to the Novelty group but familiar to the Control group. The iVR task was followed by immediate recall for 5 min and the "Igroup" presence questionnaire (Schubert, Friedmann, & Regenbrecht, 2001). On day 3, participants were asked to complete the memory task online during which they completed a remember/know/new recognition task (see Panel D). Panel B Screenshot of virtual kitchen of the iVR task with objects present as seen by the participants. Panel C Illustration of encoding task. A trial of the encoding task started with the presentation of a fixation cross (500 msec) followed by the presentation of a word (300 msec) along with reminder of current task: either deep (inanimate/animate) or (non-alphabetical/alphabetical). Participants had 4.5 sec to respond. This task had 4 blocks following an ABBA design, with A/B referring to the task. The task and stimuli were taken from Otten et al. (2001). Panel D Illustration of recognition memory task. A trial of recognition task started with a fixation cross (500 msec) followed by the presentation of the word (1000 msec) alongside a task reminder (remember/familiar/new) that stayed on screen until a response of given.

During the study phase, a trial started with a fixation cross displayed for 500 msec followed by the presentation of a word for 300 msec alongside a reminder of the current task (alphabetical vs animate). The tasks were based on Otten et al. (2001): In the alphabetical (shallow) tasks, participants decided whether or not the first and the last letter of a word are in alphabetical order, while in the animate (deep) task, they decided whether or not the presented word refers to an animate object. Participants were instructed to use one finger from each hand to press one key for non-alphabetical/inanimate words and another for alphabetical/animate words. The reminder remained on the screen for 4.5 sec comparable to Otten et al, after which the next trial started. A failure to respond in that time frame led the trial being scored as no response. The encoding task was divided into four blocks of 72 words, with each block having one of the two tasks. The order of words within a block was randomised once and then was the same for every participant. Participants were not be told that their memory for the words were later be tested; rather, they were told that their ratings of the words would simply help prepare the stimuli for another experiment on different participants.

After the study phase, participants of both groups (Novelty and Control group) immediately spent approximately 25 minutes in the iVR task from Chapter 2 Experiment 3b. Assignment of participants to groups alternated according to availability. The iVR task consisted of an encoding phase, in which participants have 45 sec to explore a virtual kitchen and memorise the locations of 20 everyday items (e.g. microwave, helmet etc.; see Figure 3.1B). After this, participants were taught how to pick up objects in this VE and then asked to pick up and place each of the twenty objects that they had seen earlier at the locations that they had to remember. After completing this, participants removed the VR headset and completed a 3AFC recognition memory task, in which participants had to choose the correct object location out of three alternatives, and a rating task, in which participants rated how expected the locations and objects were. More information on the nature of the iVR task can be found in Open Science Framework pre-registration form: https://osf.io/4sw2t/. The visuospatial nature of the iVR task (i.e. remembering object locations) was sufficiently different from memorising single words that we did not expect interference between the tasks (Wixted, 2004).

After the iVR task, participants were asked to recall as many words as they can (immediate recall) for 5 minutes. Participants were asked to write down the words that they remember on a piece of paper. After this, participants completed the Igroup Presence Questionnaire (IPQ; Schubert et al., 2001) to assess involvement, presence and realism of the iVR experience (http://www.igroup.org/pq/ipq/items.php). In a computerised adaptation of this questionnaire, participants gave their ratings on a slider scale using the original questions and anchors.

The next day, participants were asked to complete a recognition memory task similar to the one used by Fenker et al. (2008), in which the participants were presented with the words they studied previously, intermixed randomly with 144 non-studied words. A trial in this task started with a fixation cross displayed for 500 msec followed by the presentation of the word for 1000 msec. Participants were instructed to press the ‘n’ key if they think the word was new, the ‘f’ key if the word was familiar and the ‘r’ key if they remembered the word. Note that I preferred to use the word “familiar” when instructing participants because they typically found it easier to understand than the word “know”. The exact instructions were based on those used by Bastin et al. (2010), though conceptually my results were comparable to previous studies using Tulving’s original “know” instruction, which is why I continued to use the words Remember and Know in analyses below. At the end of the online session, participants were completely debriefed and compensated.

Stimuli

The same number of words (144 per task, i.e. 288 total) were used in the study phase as in Otten et al (2001), though only half of the number of unstudied words (144) were used in the test phase, since unstudied words were not of primary interest here. Thus a subset of 432 words from Otten et al (2001), balanced according to study response category (animate/inanimate and alphabetical/non-alphabetical), had been split into 3 sets of 144 words and the assignment of words to the two study conditions and the unstudied condition were counterbalanced across participants (6 different combinations in total). The lists had been selected so as not to differ in terms of the characteristics available in the MRC Psycholinguistic Database (see here for selection and list creation process; for words see Wilson, 1988).

For the iVR task, the following stimuli will be used: the virtual kitchen has been created using SketchUp (https://www.sketchup.com/), unity3d (https://unity3d.com/) and freely available 3D models downloaded from https://archive3d.net. In addition to typical kitchen furniture such chairs and a table, this kitchen contains 20 everyday objects such a hat, a calendar and a toy (for an illustration of the virtual environment see Figure 3.1B).

Equipment

At the beginning, the VE was presented with an HTC VIVE VR system and run on MSI VR ONE 7RE-057UK computer with Intel Core i7-7820HK, 16GB RAM and GeForce GTX 1070, which can be worn as backpack allowing free movement. Due to equipment failure I had to replace the VR computer and started to use a Dell Desktop PC (Precision

5820 Tower X-Series) with Intel Core i9-10900 and GeForce RTX 2080 midway through data collection. To allow free movement, I used the VIVE Wireless Adapter.

Other laboratory tasks were completed on a Dell Latitude E6530 laptop. For these tasks, stimuli were presented with Matlab (https://www.mathworks.com) using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The online tasks were run on a JATOS (Lange et al., 2015) server hosted on the MRC-CBU servers, which were compliant with data protection and security policies. That task was programmed in Javascript with jsPsych (de Leeuw, 2015).

Statistical design and hypotheses

To address the hypotheses outlined in the Introduction, I ran Bayesian *t*-tests (Morey & Rouder, 2018) using the package *BayesFactor* (version 0.9.12-4.2) for both recall and recognition tasks, with factors novelty (Novelty vs Control group, between-participants), encoding task (shallow vs deep, within-participant) and, for the recognition task, memory quality (probability of recollection vs familiarity, within-participants). The hypotheses were tested using BFs, for the alternative versus the point-null hypothesis, calculated for *t*-tests using the default scale parameter of $\sqrt{2}/2$. I used between-participant *t*-tests to test for the presence of main effects and interactions, given my directional predictions (note that, in factorial designs, within-participant factors with only two levels can be reduced to difference scores, enabling all interactions in the present design to be reduced to *t*-tests between the two groups on these difference scores; likewise, main effects can be reduced to *t*-tests between the two levels of one factor by aggregating across the other factors). Using *t*-tests for planned comparisons, instead of more the traditional ANOVA approach, enables me to accurately express my statistical hypotheses that are directed in some cases. Specifically, I predicted a main effect of novelty in recall (Hypothesis 1.1) and recognition memory (Hypothesis 1.2), with better memory in the Novelty group; an interaction between novelty and encoding task in recall (Hypothesis 2.1) and recognition (Hypothesis 2.2), with a larger novelty effect predicted for words that are shallowly encoded; an interaction between novelty and memory quality in recognition memory (3), with different probabilities of recollection and familiarity in the novelty versus control group; and the three-way interaction between novelty, encoding task and memory quality in recognition memory (4), with different probabilities of recollection and familiarity restricted to the shallowly encoded words in the novelty group. Note that the first two planned comparisons are one-tailed, while the last two are two-tailed. The directional predictions are explained in the Introduction, and I argue that directional hypotheses were justified because, according to BTT, null effects would be equivalent to negative effects and would lead to the same conclusions. In other words, that novelty could impair memory is not interesting theoretically to me other than that it will provide evidence against the BTT. This is not necessarily true for the two last comparisons because, as explained in the Introduction, different boundary conditions of BTT predict that novelty either boosts familiarity or boosts recollection.

For the recall data, the dependent variable was the number of studied words recalled. For the recognition memory data, I used a multinomial processing tree (MPT) model that is analogous to the "Source-Item" model in Cooper, Greve and Henson (2017), which assumes two underlying processes contributing to memory: the probability of recollection (*r*) and the probability of familiarity (*f*) (see Figure 3.2). In this model, recollection and familiarity are discrete states and recollection is always accompanied by familiarity. Additional parameters that are estimated but not subject to statistical test in my model are *gr* and *gk*, which are the probabilities that a guessing response leads to a Remember and Know response, respectively. Estimating these parameters effectively adjusts the estimates of recollection and familiarity by their false alarm rates. The MPT will be fit using the *"MPTinR"* package (version 1.11.0; Singmann & Kellen, 2013). For statistical analysis of the resulting parameters, probabilities were submitted to an arcsine transformation so that their values approximately follow a normal distribution and are not bound between 0 and 1. Note that I neglected to mention in my registration of this experiment the following: For comparisons that do not distinguish between recollection and familiarity I used Pr (hit – false alarm rate) instead of averaging recollection and familiarity.

In terms of choosing participant numbers, I ran a "fixed N" Bayesian analysis (Schönbrodt & Wagenmakers, 2018) for the hypotheses above. I based my sample size estimation on simulating a general linear model (GLM) with effect sizes similar to those reported in previous studies for a main effect of novelty. Based on the reported statistics, I calculated Cohen's d for four effects that in the literature. For Fenker et al. (2008) the effect size is d = 0.588 for immediate remember, d = 0.943 immediate recall and d = 0.894 for delayed remember. For Schomaker et al. (2014), the effect size for immediate recall is d = 0.873. My simulations based on the median (d = 0.884) showed that 36 participants per group is sufficient to provide compelling evidence (BF_{10} > 6) for one-tailed comparisons (1 and 2) with a probability of approximately 91%, and for two-tailed comparisons (3 and 4) with a probability of approximately 83%. At the same time, I would be able to provide compelling evidence (BF_{10} < 1/6) for the null hypothesis for the absence of an effect in 30% of the cases for one-tailed comparisons, while the probability of obtaining compelling misleading evidence would be extremely low for all comparisons ranging between 0 and 0.09% (see here for whole design analysis).

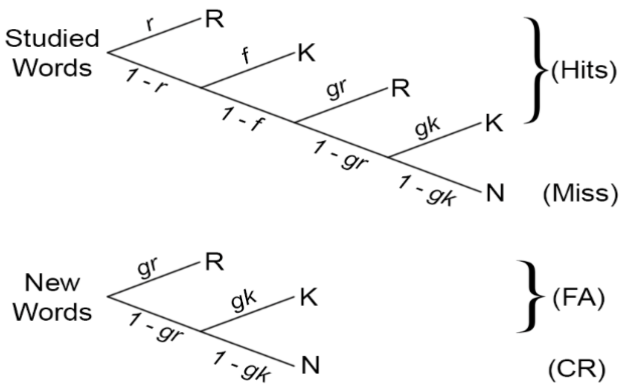

Figure 3.2: Illustration of multinomial processing tree (MPT): The model illustrated assumes two underlying processes contributing to memory. The tree at the top captures responses to studied words, while the tree at the bottom captures responses to new (unstudied) words. Possible responses in the remember/know/new task are R, K and N, respectively. In response to studied words, R and K are categorised as hits, while N responses are misses. In response to new words, R and K are categorised as false alarms (FA) and N as correct rejections (CR). Estimated MPT parameters are: r = probability of recollection, f = probability of familiarity, gr = probability of a guessing responses that leads to an R response, gk = probability of a guessing responses that leads to a K response.

3.3 Results

Note in the text I report accuracy rates and probability estimates, while the statistical tests and effect size estimates are based on arcsine transformed values, where necessary.

Encoding task

There was no compelling evidence that accuracy in the animacy condition, M = 0.884 (SD = 0.0864), was different from accuracy in the alphabetical condition, M = 0.897 (SD = 0.101), BF_{01} = 2.51, d = 0.183.

However, as expected, responses were faster in the animacy condition, M = 1000 msec (SD = 227 msec), compared to the alphabetical condition, M = 1540 msec (SD = 351 msec), BF_{10} = 2.51e+21, *d* = 1.88 (therefore any better retrieval in the animacy condition is unlikely to simply reflect a longer time spent studying).

Memory tasks

Despite not explicitly registering in Stage 1, I first report the “levels of processing” effect in order to confirm that this manipulation had the desired effect of producing greater memory for deeply- than shallowly-encoded words. For the immediate recall, participants recalled more words in the animacy condition, M = 15.4 (SD = 6.38), than in the alphabetical condition, M = 2.77 (SD = 2.85), BF_{10} = 8.05e+27, *d* = 1.91. Thus my level of processing manipulation had the intended effect. Full results for recall and recognition task can be found in Figure 3.3.

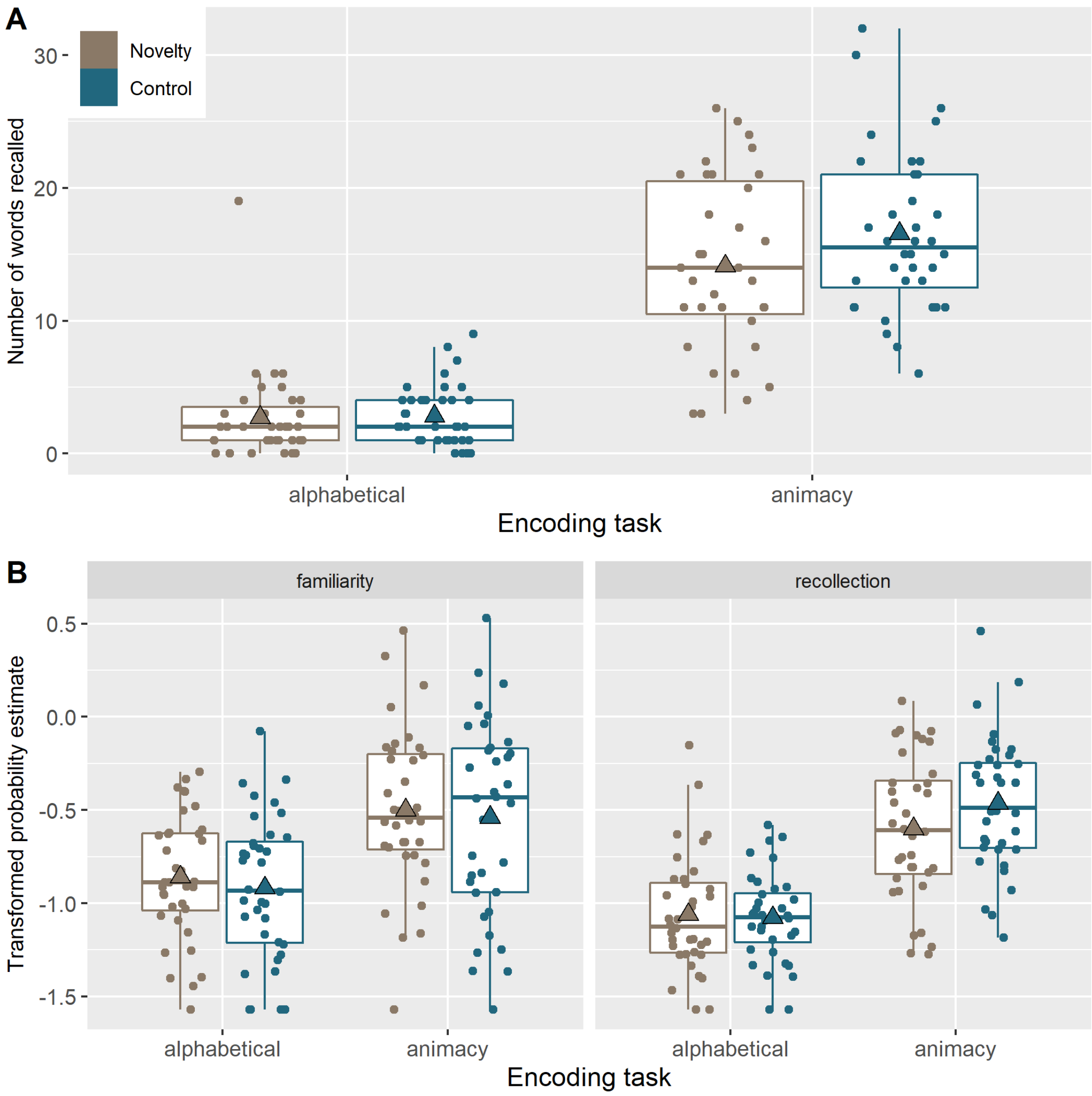

Figure 3.3: Main results for recall and recognition: Panel A Number of words recalled as a function of the encoding task alphabetical/non-alphabetical and animate/inanimate for the Novelty group (brown) and the Control group (blue). Panel B Arcsine transformed probability estimates from the MTP analysis of the recognition task. Estimates for familiarity (left) and recollection (right) are displayed as function of encoding task for the Novelty and the Control group. Boxplots with triangles illustrating the mean.

For registered Hypothesis 1.1 (Figure 3.4A), the immediate recall tests provided evidence that participants in the Novelty group, M = 16.8 (SD = 7.02) did not recall more words than participants in the Control group, M = 19.4 (SD = 7.59), i.e., moderate evidence for the null hypothesis, BF_{01} = 9.2, d = 0.35. Based on this I stopped data collection. In a two-tailed version of this test (not registered), evidence that the Control group actually recalled more words than the Novelty group was inconclusive, BF_{10} = 0.618, despite the difference in means.

For Hypothesis 1.2 (Figure 3.4B), an overall measure of Pr for the delayed recognition memory test provided moderate evidence that the Novelty group, M = 0.196 (SD = 0.091), did not have better memory than the Control group, M = 0.22 (SD = 0.106), BF_{01} = 7.46, d = 0.25.

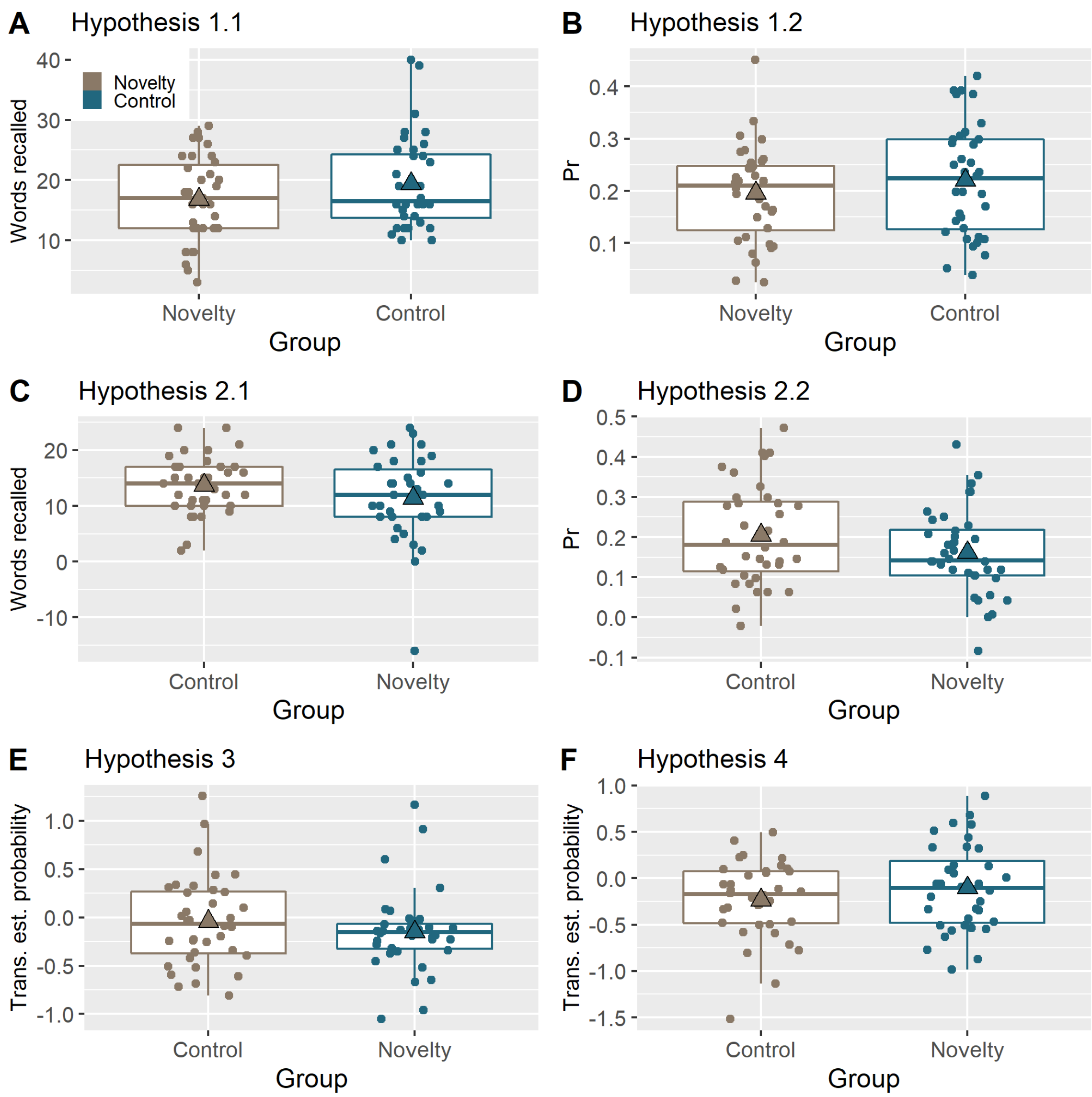

Figure 3.4: Group difference plots: Boxplots for each hypothesis showing the values or difference scores that are compared with the Bayesian *t*-tests with triangles illustrating the mean. The novelty group is displayed in brown and the control group in blue. The y-axes for Panel E and F are short for transformed estimated probability.

For Hypothesis 2, I found inconclusive evidence that the levels of processing effect for immediate recall differed between the Novelty group, M = 11.4 (SD = 7.71), and Control group, M = 13.7 (SD = 5.1), BF_{10} = 1.17, d = 0.356 (Hypothesis 2.1; Figure 3.4C), as that it differed for delayed recognition between the Novelty group, M = 0.162 (SD = 0.108), and the Control group, M = 0.205 (SD = 0.125), BF_{10} = 1.29, d = 0.371 (Hypothesis 2.2; Figure 3.4D).

For Hypothesis 3 (Figure 3.4E), I found inconclusive evidence (two-tailed) of a difference in recollection versus familiarity during delayed recognition (i.e., difference in the "r" and "f" parameters from the MPT, collapsed across two encoding conditions) between the Novelty group, M = -0.0509 (SD = 0.152), and Control group, M = -0.024 (SD = 0.164), BF_{10} = 0.385, d = 0.244.

To test for the interaction of Hypothesis 4 (Figure 3.4F), I first calculated the difference between the two encoding conditions (deep versus shallow) for each MPT parameter (r versus f), then subtracted these difference scores and compared them across groups. Again, there was inconclusive evidence for any difference between the Novelty group, M = -0.0162 (SD = 0.195), and the Control group, M = -0.0715 (SD = 0.186), BF_{10} = 0.486, d = 0.3. Table 3.1 shows a summary of BF for each Hypothesis.

Table 3.1: A summary of the results of the six registered hypotheses.

Hypothesis	Description	BF_{10}	BF_{01}
1.1	Main effect of novelty in immediate recall	0.109	9.200
1.2	Main effect of novelty in delayed recognition	0.134	7.460
2.1	Interaction between novelty and encoding task in recall	1.170	0.855
2.2	Interaction between novelty and encoding task in recognition	1.290	0.774
3	Interaction between novelty and memory quality in recognition	0.385	2.600
4	Three-way interaction between novelty, encoding task and memory quality in recognition	0.486	2.060

Post VR questionnaire

As an additional non-registered analysis, I examined the data from the post VR questionnaire that I collected. For this, the data was rescaled to vary from 0 to 6, as in the original scale (Schubert et al., 2001), with inverse items reversed[7]. The IPQ score was then calculated by summing across all items (similar to Schomaker et al., 2014, as confirmed by personal communication). This analysis showed that the Novelty group, M = 51.8 (SD = 10.9), did not differ from the Control group, M = 52 (SD = 10.1), d = 0.0238, BF_{10} = 0.246 (Figure 3.5A). In addition, I asked participants to rate the statements: “This experience was novel”, “This experience was exciting” and “This experience was uncomfortable”. However group differences did not arise for any of these statements: most surprisingly for Statement 1, concerning subjective rating of novelty, there was anecdotal evidence against the Novelty group, M = 5.28 (SD = 0.957), finding the iVR experience more novel than the Control group, M = 5.18 (SD = 1.34), d = 0.049, BF_{10} = 0.249 (Figure 3.5B). Note to deal with the negative skew for Statement 1, data were scaled to be from 0 to 1 and then transformed with arcsine transformation for the analysis only.

[7] http://www.igroup.org/pq/ipq/data.php

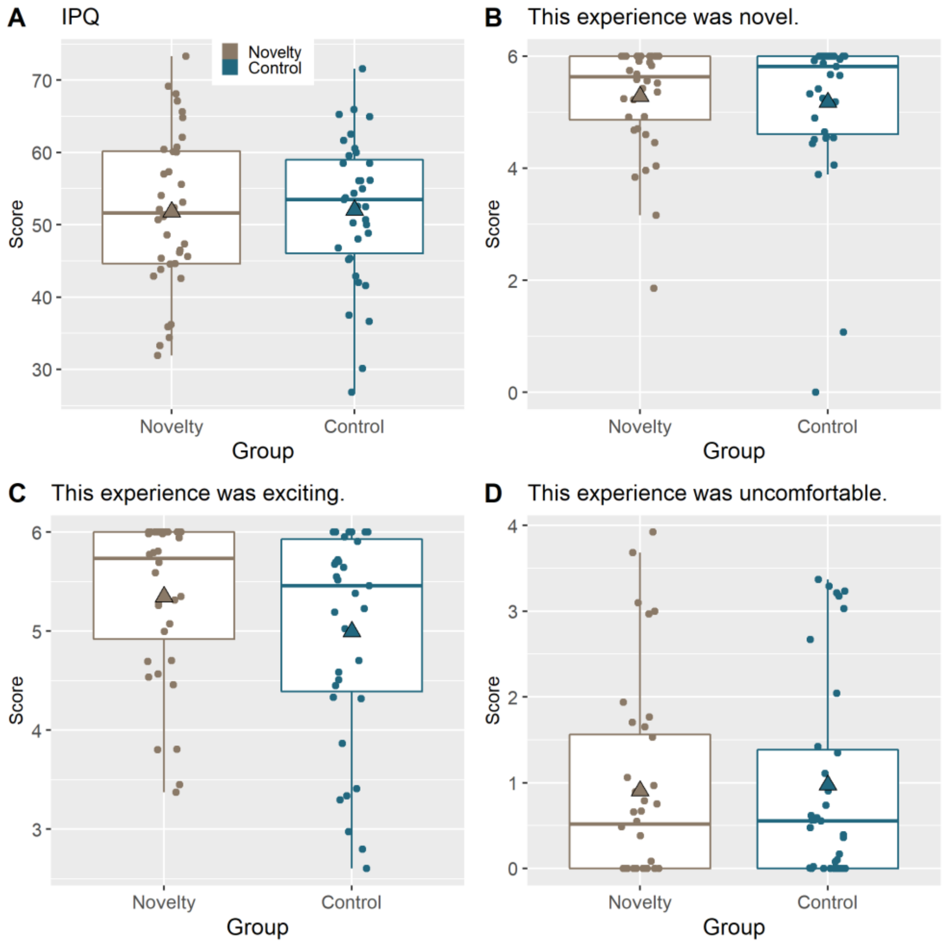

Figure 3.5: Boxplots for questionnaire data, with triangles showing the mean. The novelty group is displayed in brown and the control group in blue.

3.4 Discussion

This study found no evidence that a novel experience can retroactively enhance memory for material that had been learned prior to that experience. For the novel experience, I used people's first experience of an iVR environment (compared to other people's second experience); for the material to be remembered, I used a list of unrelated words, which were studied under incidental tasks of either animate/inanimate judgment (deep encoding) or alphabetical judgment (shallow encoding). BFs showed moderate evidence that memory was not enhanced for same-day recall of the words (immediately after the iVR experience), and moderate evidence for no effect on recognition memory tested the next day. I also found no evidence (either way) for my additional hypotheses that any novelty-related boost would be greater for shallowly- than deeply-encoded words, and would differ for words later recollected versus familiar in the delayed recognition test.

The lack of any effects is surprising because previous studies did find that experiencing a novel VE versus a familiar one can boost memory for words learned *after* the VE experience (Schomaker et al., 2014, 2021). One reason could be that this type of novelty only has proactive effects on memory for this type of material, and no retroactive effects. However, this is contrary to animal studies that tend to find both proactive and retroactive effects (but with different memory tasks; see Section 3.1 of Chapter 1) and the BTT derived from these animal data.

Another possibility is that my comparison of first versus second experience of iVR did not differ sufficiently in terms of novelty. This might explain why there was no evidence for a difference between my two groups in their mean rating of novelty (or any other aspect of their experience) in my post-experiment questionnaire, though the lack of difference in subjective ratings of novelty could be obscured by a ceiling effect. I chose this comparison of first versus second experience of a VE because it was tightly controlled, compared for example to comparing an iVR experience to a more familiar, non-iVR experience, which could differ in ways other than novelty, and because a similar comparison was used by Schomaker et al. (2014; 2021). In fact, Schomaker et al. compared a novel VE with a familiar VE, for participants who were generally familiar with the technology, which if anything would seem a less extreme contrast in novelty than my comparison of participants' first ever experience of iVR: i.e. my groups differed in their familiarity for both the VE (my virtual kitchen) and the iVR technology (since I excluded participants with previous iVR experience). The claim that my comparison was not novel enough also seems to conflict with previous human studies that have simply used familiar versus novel static images on a computer screen (e.g., Fenker et al., 2008).

The lack of a difference between groups in the score given on my post-experiment novelty question could also have other reasons. Foremost, participants in the two groups may have used different references for their novelty rating (despite experiencing different levels of absolute novelty), e.g., participants in the Control group might have rated the novelty of their second iVR experience relative to other experiences that day (Day 2), rather than explicitly refer back to their first iVR experience on the previous day (Day 1); or they may have misunderstood the question, and rated their novelty for the overall experiment over the two days. It is also worth noting that most previous studies did not report subjective experiences of novelty for their manipulations. Schomaker et al (2014) used the Igroup IPQ questionnaire (Schubert et al., 2001), which is commonly used to measure presence, involvement and realism in VR experiments but not novelty *per se*. While presence ratings were higher in Schomaker et al (2014) after being in a novel versus a familiar VE, no difference as a function of novelty was found in Schomaker et al. (2021) like here.

Another possible reason for the lack of a novelty-boost on memory is that the boost was masked by the fact that there were differences between the Novelty and the Control group in terms of the difficulty of the task both groups completed between word learning and the memory test (i.e., count the number of objects in the kitchen, and then replace objects at their previous location, as part of the experiment in Chapter 2). It has been claimed that demanding activities can impair consolidation of memories (Dewar et al., 2007; Wixted, 2004). Indeed, my task was likely to be easier the second time it was performed (in the Control group), which might have resulted in less impairment of consolidation than in the Novelty group, counter-acting any advantage of novelty. While this is possible, I note that in most situations, including in real-life, novelty is generally associated with greater cognitive demands (to process the novelty), so this potential confound would appear to apply previous demonstrations too (such as novel lessons in children's schooling, Ballarini et al., 2013; Ramirez Butavand et al., 2020).

While I could not provide evidence for behavioural tagging with my experiment, there is still considerable evidence that, in some situations, effects of a novel or surprising experience can affect memory for unrelated preceding or succeeding information, highlighting an important limitation of PIMMS. Memory encoding is therefore not just driven by PE, but also by processes that happen around it, and to understand this, neurobiological and computational accounts must be fused (see Chapter 5).

4 SETTING BOUNDARIES

4.1 Introduction

Besides boosting memory, surprise also plays an important role in parsing and segmenting our continuous experience. As discussed in Chapter 1, theories like EST or EHM explain how boundaries triggered by experiencing shifts that typically lead to PE affect memory performance in several different ways. In this chapter, I focus on the effects of spatial boundaries (i.e. doors) on temporal order memory.

There is a range of studies that show that memory for the temporal order of two items is better when those items belong to the same event (DuBrow & Davachi, 2013, 2016; Ezzyat & Davachi, 2014; Heusser et al., 2018; Horner et al., 2016). For instance, DuBrow and Davachi (2013, Experiment 1) presented a sequence of faces or objects, with the category changing every few trials, allowing the authors to examine the effect of the changing category on temporal order memory. Associative processing was emphasised by instructing participants to create associations between images either through narrative construction or vivid mental imagery. The main way temporal order memory was assessed was to present pairs of items from the same category that either contained or did not contain a boundary between them, and asking which of the two was more recent. The main result was that temporal order memory was better when no switch (i.e. boundary) occurred.

As already mentioned in Chapter 1, a special type of boundary is walking through a doorway. Horner et al. (2016) developed a spatial paradigm conceptually similar to the one used in DuBrow and Davachi (2013), except that participants navigated through a series of rooms, each of which contained two tables. Placed on those tables was an image of an object for which participants had to decide whether it was natural or man-made. This paradigm allowed the authors to investigate how temporal order memory is affected as a function of whether two objects appeared in the same room (within boundaries) or in two adjacent rooms (across boundaries). Temporal order memory was tested in slightly different way than is typically done (see e.g. DuBrow and Davachi, 2013), by presenting a cue object and asking either "Which object came next?" or "Which object came before?". Here, Horner et al. offered their participants three alternatives (i.e. a 3AFC task) with one correct and two incorrect response options (one seen earlier in the sequence and the other one seen later in the sequence). Across three experiments, temporal order memory was found to be better for objects within the same room vs. across two rooms. In order to compare the results from my experiments, I downloaded Horner's et al.'s data and plotted them in Figure 4.3. The authors also tested item memory, whether the first of the second object of the room would be remembered differently and what they coined context memory (i.e. the exact room in which an object was presented). Item memory, though above chance, did not vary as a function of whether the object was presented first or second in a room, while context memory was consistently not above chance.

Thus it appears that walking into a new room can act like a boundary in its effect on temporal order memory. However, it remains unclear what aspect of walking through a doorway is driving this and other related memory effects (e.g. the location updating effect of Radvansky & Copeland, 2006, described in Chapter 1). One possibility could be that doors are special and always serve as boundaries regardless of any other factor. However, walking between rooms also often results in a large perceptual change (PC). For instance, real world rooms typically differ in terms of their layout and what you find in them (e.g. a bathroom versus a living room). On the other hand, EST posits that the PE is central for setting boundaries (e.g. a PC may not be important if it is fully expected when walking between rooms). The original aim of the work in this chapter was therefore to examine whether PC and/or PE are responsible for the boundary effect of walking through a doorway in a virtual environment.

To study effect of PC and PE, I created a (non-immersive) VR paradigm that closely matches the experiments described in Horner et al. (2016). To expand this paradigm, I designed a special "M-room" (Figure 4.1B), as distinct from the "open" or "O-rooms" like those used by Horner et al. (2016; a video showing the difference between O-rooms and M-rooms can be found here https://vimeo.com/532276947). Participants watched a video from the perspective of someone traversing these rooms in a fixed route (participants did not themselves navigate; see below). When traversing an M-room, the viewer can only see one half of the room until they reach the middle section, at which point their perspective is rotated and they see the other half of the room. The middle section of the M-room was kept neutral white. Approaching this neutral area is depicted in the first two images of Figure 4.1 Panel C. The fact that the viewer does not see the whole room when they enter allows certain features of the second room half (e.g. wall colour or floor pattern) to be changed, i.e. to manipulate PC, in the absence of a door between the two parts of the room. PC, or the absence of PC, could then be cued, thereby manipulating PE by building up and confirming/violating the expectation of PC. For instance, a cue could indicate that the wall colour of the second room will change to green, while it actually would stay the same colour as the first half. This example would be a case where there is no PC but PE, but of course other combinations are possible. Memory for objects that straddled this PC/PE in an M-room could be compared with that in the more standard "across room" condition (as well as memory in a "within room" condition, in which neither PC nor PE are present). If introducing PC and/or PE rendered performance similar to the "across room" condition, I could infer that the specific combinations of factors is driving the standard effect of walking through a doorway.

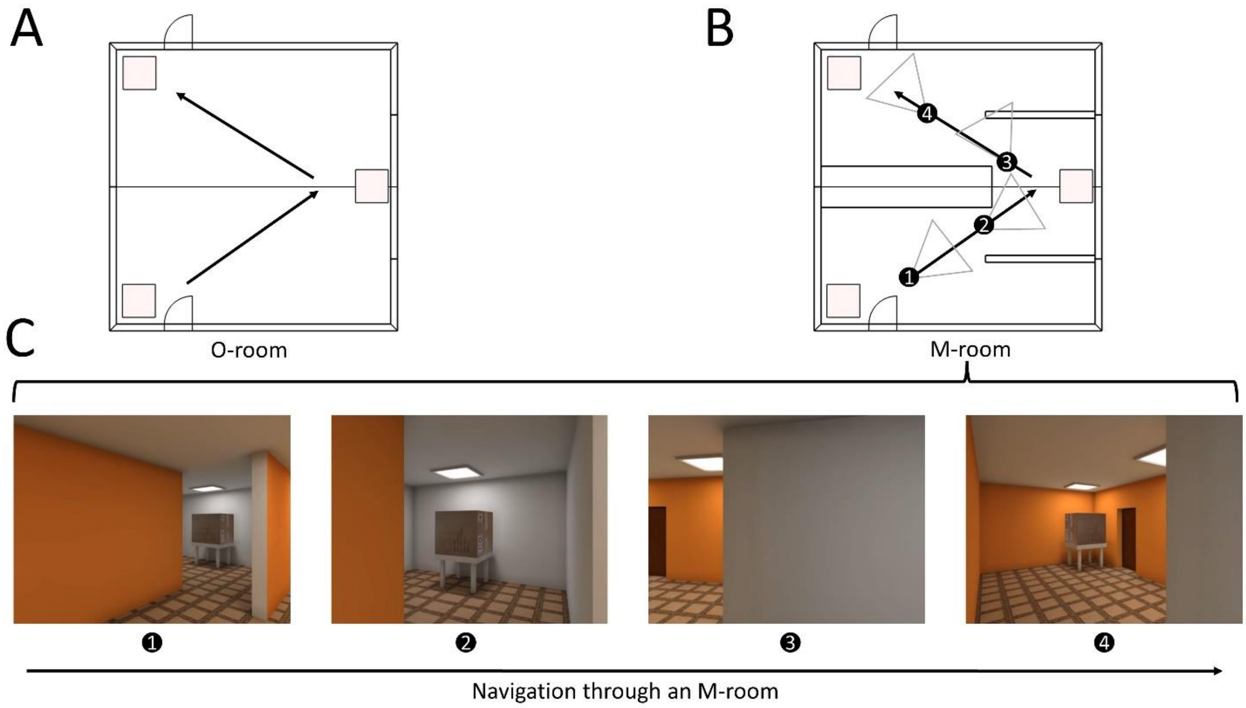

Figure 4.1: Illustration of room layout: Panel A shows the layout of the standard O-room. Panel B shows the layout of the M-room with arrows that illustrate the navigation path. The numbered triangles approximate the camera position and orientation of the screenshots shown below. Panel C shows a series of screenshots that illustrate the camera view while navigating through an M-room. Note that in Panel C, no PC is introduced as both halves of the room have the same wall colour. For further illustration, see the video (https://vimeo.com/532276947), which illustrates the difference between watching navigation through an M vs. O-room.

As a first step in this project, a pilot study was conducted to verify that crossing to the second half of the M-room in the absence of PC or PE does not constitute a boundary in itself. To this end, I tested whether the superior temporal order memory (i.e. which object appeared before/after a cue object, as in Horner et al. 2016; see below) for objects within the same room than in different rooms, is similar in M-rooms and O-rooms.

Unfortunately, this basic effect of superior temporal order memory for objects within versus across rooms could not be replicated in either room type. Therefore, the planned manipulation of PE versus PC was not performed, and the three experiments that follow only demonstrate an inability to replicate the basic room-boundary effect (i.e. within > across), even after addressing several potential confounds. Remaining differences between the present experiments and those of Horner et al. (2016) are discussed, as fruitful avenues for future experiments.

4.2 Experiment 1

Method

Procedure

While in Horner et al. (2016) participants navigated through the VE themselves, participants in this series of experiments watched a video from a first-person perspective of someone walking through the rooms. This decision was made to ease online data collection (which was enforced in part by the pandemic that occurred during the third and fourth years of this PhD). While watching the video, participants in this task were required to judge whether an object was smaller or bigger than a reference (a shoebox like in DuBrow & Davachi, 2013) as soon as the object appeared. In this and all subsequent experiments, the object was visible for 3 sec. After the 3 sec, the object disappeared and the cardboard box re-appeared controlling the time during which the object was visible for the participant. Participants were instructed to press S on their keyboard if the object is smaller and L if the object is larger than the reference object. Key presses were recorded but not further analysed.

After watching the video, participants immediately completed some memory tasks. The first memory task was temporal order memory, assessed by a 3AFC task. In this task, a cue object was presented along with the question “What came before this object?”, or “What came after this object?”, depending on the counterbalance condition. Beneath the cue object, three objects were presented: one correct (i.e. the target) and two foils. This was followed by two further memory tests in which participants were first asked to choose in which type of room the cue object was seen in (in-house piloting showed that people could not remember the exact room – in terms of wall colour and floor texture – hence only the type of room was tested similarly to the Horner study) and then asked on which of the two tables the cue object was presented (i.e. 2AFC task in both cases). In contrast to Horner et al., (2016), I did not use a context memory task (apart from pre-pilot where memory performance was not above chance either) and I also decided against implementing an item memory task as finding suitable 3D objects is cumbersome and no interesting results with this test were reported in Horner et al. (2016). Finally there was a short debrief questionnaire, which was mostly used to screen for technical difficulty and participant feedback.

All tasks in this chapter were run online on a JATOS (Lange et al., 2015) server hosted on the MRC-CBU servers, which were compliant with data protection and security policies. That task was programmed in Javascript with jsPsych (de Leeuw, 2015).

Virtual environment and objects

The basic layout of both rooms (Figure 4.1) were created with SketchUp (https://www.sketchup.com). These were then imported into the video game development platform unity3d (https://unity.com/), which was used for creating the VE including the necessary programming, lighting and camera settings. In contrast to Horner et al., the layouts of all O and M-rooms were always identical including the positions of the tables, differing only in wall colour and floor material (wood or carpet textures). Like the order of the objects, the wall colours and floor materials were constant across the videos of the same counterbalance condition (see below). The number of wall colours (blue, brown, green, grey, orange, pink, purple, red, turquoise and yellow) and floor materials (5 different carpets and 5 different wood floors) allowed me to construct 45 unique rooms. Both types of rooms contained three tables; however only two were used in the experiments discussed here (the remaining table was for a possible future version).

88 everyday objects were downloaded (e.g. guitar, toys, household items etc.) from archive3d (https://archive3d.net/) and edited either in blender (https://www.blender.org/) or in unity3d itself. I tried to find as many 3D versions of objects as possible that were also used in Horner et al. (2016). The objects were presented at realistic scales (i.e., an

object that is small in reality was also small in the virtual environment) to avoid any PE stemming from unrealistic object sizes.

Videos were simply created by capturing the screen while I navigated through the series of room using keyboard and mouse controls. Different conditions entailed different videos. Great care was taken to minimise any difference between the videos, though it is possible that some differences existed.

Counterbalancing

In Experiment 1, M and O-rooms alternated so that each participant saw the both types of rooms. Four videos were created where the order of the objects presented was the same, but two factors were counter-balanced: a) whether the first association between objects was across vs. within rooms and, b) whether the first room the participants encountered was an O or an M-room. Thus, two videos (Video 1 and 2) were created where two objects were presented in the first room (so the first object-object association was within-room), while in the other two (Video 3 and 4), the first room only showed one object (so the first object-object association was across-rooms). To control any influence of which room type (O or M), the first room in Video 1 and 3 was an M-room, while the first room in Video 2 and 4 was an O-room. An additional between-subject condition was the direction of the temporal order question (before vs. after), producing a total of eight counter-balance conditions (which were not analysed).

Foil selection and trial design

In order to implement tighter control of the foils' temporal position than it was done in Horner et al. (2016), where the foils were selected randomly (e.g., some could have come from the start or end of the experiment), foil 1 was the object that was presented four positions before the target object and foil 2 was the object that was presented four positions after the target (see Figure 4.2). The foils were therefore presented in the same room type and on the same table as the target object.

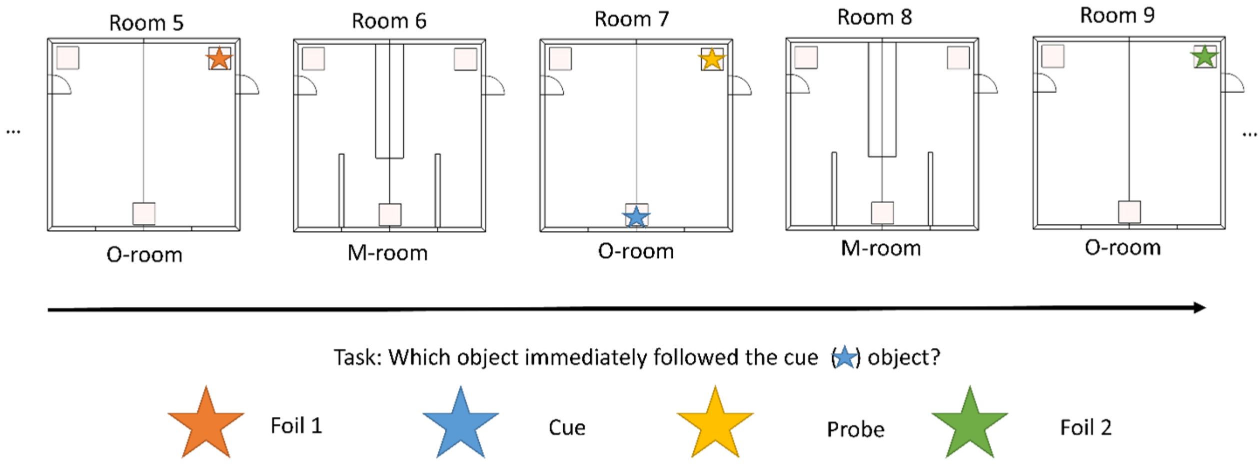

Figure 4.2: Illustration of foils selection in Experiment 1a and 1b: The cue object (blue star) is followed by the probe (yellow star). Foil 1 (red star) and foil 2 (green star) are found three/four positions away from the probe but are in the same room type and on the same table as the probe. Illustrated here is the forward question.

Each object only served as cue once. Due to the availability of suitable foils, the number of associations tested was lower than the total number of objects presented in the videos. For across-room associations there were 49 trials. Dependent on the counterbalance condition, there were either 19 trials for within-M-room associations and 20 trials for within-O-room associations, or the other way around.

Experiment 1a

Sample

In this first "pilot" experiment, 10 participants (6 female and 4 male) were tested through Prolific (even though not fully counterbalanced), with mean age M = 33.3 (SD = 9.24) years. In this and all other experiments, participants using mobile devices were not able to participate to make sure that they used a laptop or desktop PC.

Statistical analysis

Bayesian t-tests were used to compare mean differences. To this end, *ttestBF()* from the BayesFactor package (Morey & Rouder, 2018; version 0.9.12-4.2) was used[8]. In addition, Bayesian ANOVAs were calculate with *anovaBF*(). The parameter rscale was set to the default value of $\sqrt{2}/2$ or *t*-tests and 1/2 for ANOVAs. Based on previous work (e.g. Horner et al., 2016), it was expected that performance in the "within" boundary condition would be better than in the "across" boundary condition (i.e. higher proportion of correct responses). Therefore, directed (i.e. one-sided) tests were used where applicable (for instance tests against chance were one-sided). Proportion correct data were arcsine-transformed to create more normally-distributed data for analysis but original values are reported in the text.

Note that, because of the alternation of the two types of room, the preceding room in the "across" room condition was always of the opposite type to the room containing the cue object. Nonetheless, a trial could be classified as an O-room or M-room depending on which room the cue object was in. The data were therefore analysed as a 2 x 2 factorial design with: "room-type" (O or M) and "boundary" (i.e., within same room vs between previous/subsequent room of opposite type).

Data and code for all three experiments can be found in this GitHub repository https://github.com/JAQuent/boundaryVR.

Results

The results for the test of temporal order memory for each condition are shown in Figure 4.3D. The 2 x 2 repeated-measure ANOVA showed anecdotal evidence against a main

[8] One might ask why I did not use a similar sequential analysis as I did in Chapter 2 and propagated the evidence. With *brms*, propagation of evidence is only suitable if I had used the same model for essentially the same task as was the case in Chapter 2. Even within Experiment 1, where the task structure is very similar, it can be argued that with chance performance in Experiment 1a & 1b it is problematic to combine these experiments with Experiment 1c that is the only experiment where performance is above chance. Furthermore, Experiment 2 & 3 do not share the same main model and also include important differences in design (blocked question type vs. two study-test cycles). For these reasons, propagation of evidence is not suitable way to analyse data here.

effect of boundary, BF_{01} = 3.14. There was inconclusive evidence for a main effect of room-type of cue, BF_{01} = 2.26, as well as for the interaction, BF_{01} = 1.49. For completeness, simple effects showed anecdotal evidence against a boundary effect for either M-rooms, BF_{01} = 3.11, d = -0.0179, and inconclusive evidence for O-rooms, BF_{01} = 2.7, d = 0.0757.

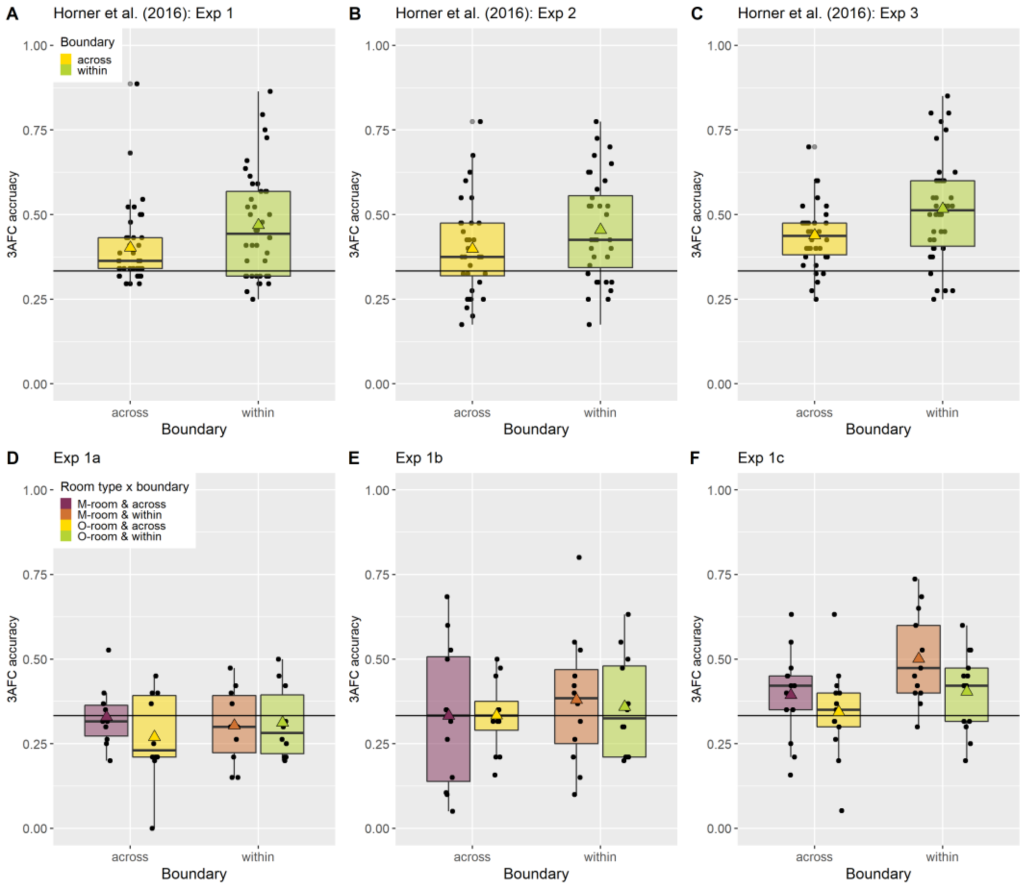

Figure 4.3: Data from Horner et al. (2016) and from Experiment 1a, 1b and 1c: Top row: The data is from Horner et al. (2016) plotted for comparison. Data were aggregated across the different question types where applicable. For Experiment 2 & 3, the data are from the original analysis, not the control analysis that tried to address difference in navigation time (see Horner et al., 2016 for details). Note that Horner et al. (2016) Experiment 1 was most similar to Experiment 1c, while Experiment 3 here was most similar to Experiment 3 in their paper. The colours green and yellow were chosen because the rooms in Horner et al. (2016) most closely matched O-rooms. Triangles represent condition mean. The data were accessed from https://figshare.com/articles/dataset/Excel_file_for_behavioural_data_presented_in_paper/1609803/3. Bottom row: Main results for Exp 1a, 1b and 1c. Standard

coloured boxplots with triangles illustrating respective condition mean for the interaction of room type (M-room vs. O-room) and boundary (across vs. within). Horizontal line indicating chance performance (1/3) for 3AFC accuracy.

However, the lack of condition differences could be because overall performance was at floor. Indeed, there was moderate evidence that temporal order memory was not above chance (i.e., averaged across within and across conditions, BF_{01} = 6.75, M = 0.304 (SD = 0.0628)).

For the tests of source memory (Figure 4.4A), there was anecdotal evidence that participants could not remember in the room-type in which an object was presented, BF_{01} = 5.07, M = 0.482 (SD = 0.0755). There was inconclusive evidence about whether or not they could remember the table on which an object was presented, BF_{10} = 1.29, M = 0.536 (SD = 0.0769).

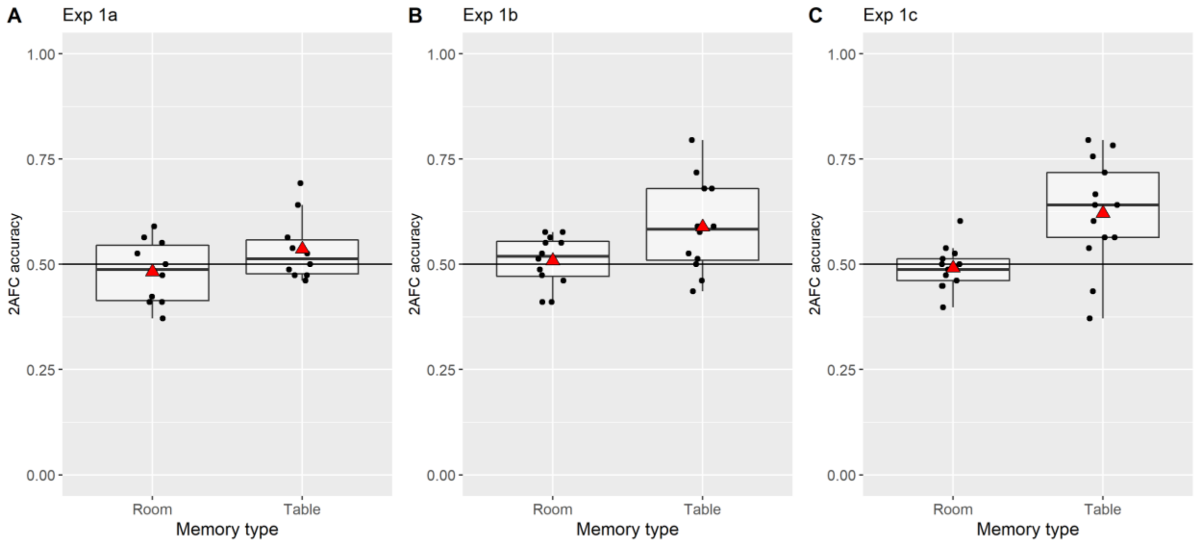

Figure 4.4: Performance in the room & table task for Experiment 1a, 1b and 1c: Standard boxplots with triangles illustrating task mean. Horizontal line indicating chance performance (0.5) for 2AFC accuracy.

Discussion

Contrary to expectations, there was no boundary effect (better temporal order memory for within versus between rooms) for either room type: the original O-rooms like those used by Horner et al (2016), nor the new "M-rooms" (despite no PC between the two tables). This may be because temporal order memory was not above chance. One reason for this could be that the task instructions were ambiguous, in that the question "What came before this object?" could be interpreted in a way that participants thought that either of the two objects that appeared before the cue object (i.e. the target as well as the earlier foil) would be a valid answer. Therefore, in Experiment 1b, the instruction was changed.

Experiment 1b

Experiment 1b was identical to 1a apart from the temporal order task instructions: Participants were now asked “In the video you just watched, which one of the three objects at the bottom of the screen appeared immediately after this object?” in order to make sure that there was no misunderstanding about which was the correct object (and also only the “after” direction of the temporal order question was tested).

Sample

In total, 12 participants, recruited through Prolific (five female and seven male), completed this version of the experiment, with mean age of M = 32.8 (SD = 11.0) years.

Results

The 2 x 2 ANOVA on temporal order showed anecdotal evidence against a main effect of boundary, BF_{01} = 3.11, inconclusive evidence against a main effect of room type, BF_{01} = 1.27, and anecdotal evidence against an interaction, BF_{01} = 3.06 (see Figure 4.3E). No boundary effect was found for M-rooms for accuracy, across, BF_{01} = 1.59, *d* = 0.256, and for O-rooms for accuracy, BF_{01} = 2.52, *d* = 0.117. However, there was also inconclusive evidence that overall performance was above chance, BF_{01} = 2.23, M = 0.351 (SD = 0.104).

For source memory (Figure 4.4B), there was inconclusive evidence that memory for the room-type of the cue object was above chance, BF_{01} = 2.33, M = 0.509 (SD = 0.06), but (unlike Experiment 1a), there was now moderate evidence that participants could remember on which table type a cue object was presented, BF_{10} = 7.16, M = 0.589 (SD = 0.11).

Discussion

Temporal order memory was still not above chance, suggesting that the previous instructions were not a problem. However, this chance-level performance might also be why there was no evidence for a boundary effect on memory. Once possible reason for such poor performance could be that the foils were very close (in time) to the target, whereas the random selection of foils in Horner et al. (2016) would tend to include foils that were much further away in time, rendering the task easier (other possible differences between the studies are considered in the Discussion).

Experiment 1c

Experiment 1c used the same (random) foil selection procedure as Horner et al (2016), with the hope that this would improve memory performance. The foils here are selected randomly apart from three rules: First, foils could not be in the same room with target. Second, foils could not be from adjacent rooms. Third, foils could be repeatedly sampled. As a result, dependent on the counterbalance condition, the average distance between the Target and Foil 1 was either -24.2 or -21.4 intervening objects, and the average distance between the Target and Foil 2 was either 21.3 or 25.2 (c.f. the distance of 4 used in Experiments 1a-1b).

Sample

Due to technical problems, some Prolific submissions were not usable. Unfortunately, the data containing the Prolific IDs were deleted so that it not possible to retrieve the demographical data for the final sample of 13 participants. However, these came from a population of 31 (10 female and 21 male), with mean age of M = 30.6 (SD = 7.97) years, who clicked on the task. Of those, majority of those immediately aborted the task for unknown reasons.

Results

This time, there was strong evidence that overall temporal order memory was above chance, BF_{10} = 24.5, M = 0.41 (SD = 0.0778); see Figure 4.3F. However, the 2 x 2 ANOVA still showed inconclusive evidence for a main effect of boundary, BF_{01} = 0.561[9].

[9] Based on the knowledge of classical statistics, one might expect the BF for the main effect in the ANOVA (BF_{10} = 1.78) should be similar to the BF that was calculated for the t-test where the data were collapsed across room type (BF_{10} = 6.77). However, models in the *BayesFactor* package are linked through the estimated value of the error variance (personal communication with Richard Morey; also see Rouder et al., 2012). This means that the factors are not independent in the same way that is the case in a classical ANOVA. I therefore also reported the *t*-test as only this fully ignores the effect of room type.

There was anecdotal evidence against a main effect of room type, BF_{01} = 3.94, and no evidence for an interaction, BF_{01} = 1.47.

Interestingly, the difference between condition (across vs. within) collapsed across room type was large, across: M = 0.369 (SD = 0.0962), within: M = 0.452 (SD = 0.0847), BF_{10} = 6.77, d = 0.876. However, despite the lack of evidence for an interaction, a boundary effect was found for M-rooms [across: M = 0.369 (SD = 0.0962), within: M = 0.501 (SD = 0.132), BF_{10} = 7.69, *d* = 0.774], even if none was found for O-rooms, across [M = 0.369 (SD = 0.0962), within: M = 0.404 (SD = 0.119), BF_{01} = 0.754, *d* = 0.43].

In terms of source memory (Figure 4.4C), there was inconclusive evidence of above-chance performance for remembering in which room type a cue object was presented, BF_{01} = 2.15, M = 0.51 (SD = 0.0583), but there was strong evidence (similar to Experiment 1b) that participants could remember on which table type a cue object was presented, BF_{10} = 19.3, M = 0.621 (SD = 0.128).

Discussion

Relaxing the selection of foils did have the intended effect of improving overall memory performance for temporal order. Surprisingly however, a boundary effect was only found for M-rooms; not O-rooms. Above-chance memory performance for the temporal order (and the table question) suggests the lack of a boundary effect does not simply owe to reduced participant engagement, though it is possible that higher overall levels of performance (more comparable to Horner et al, 2016) are needed to detect a boundary effect.

The result is disappointing because the purpose of Experiment 1 was to ascertain that walking through the middle of an M-room does not have the same boundary effect that was reported in Horner et al. (2016) for walking through doors with ‘normal’ (O) rooms (despite the fact that the boundary effect was numerically larger in the M-room). One possibility is that there is some type of interference caused by alternating between the two types of room, so in Experiment 2, room-type was manipulated between participants rather than alternating.

4.3 Experiment 2

Relative to Experiment 1, the most important change was that room-type was now a between-participant factor rather than alternating within the same video. Furthermore, participants were asked both directions of temporal order (before vs. after) in blocks instead of just one (before).

Within each room-type, the first inter-item connection (within vs. across) and the direction of temporal order memory tested first (before vs. after) were counterbalanced across every fourth participant. To prevent the experiment from getting too long, the previous room-type and table-type questions were dropped, because there was sufficient evidence from Experiments 1a-1c that participants could only remember the table on which an object was placed. Memory for the table was even present when temporal order memory was barely above chance indicating that participants tried to perform well even in Experiment 1a & 1b.

Small changes were made to the lighting in the rooms, to render them more realistic. Furthermore, camera movement was now scripted instead of generated by recording the screen while I navigated through the rooms. This eliminated any remaining between-video differences. As a consequence, the time it took to move between the object-viewing positions within a room versus between rooms was further minimised (approximately 7.8 vs 7.9 sec respectively). Other changes were that some objects were edited (colour and size) and some were entirely swapped because the debriefing questionnaire indicated that those were difficult to recognise.

Other procedural details including the foil selection were the same as in Experiment 1c. Due the availability of foils, the number of trials per counterbalance condition was either 80 and 82 for each question type (i.e. 160 and 164 in total).

Sequential data collection

While Experiments 1a-1c were pilots, Experiment 2 aimed to be properly powered. Since Experiment 1c showed a boundary effect for M-rooms but not for O-rooms, finding a boundary effect for O-rooms was the focus for Experiment 2. Data were collected sequentially for Experiment 2 for efficiency reasons. The data collection plan[10] follows:

[10] The plan was not formally pre-registered but agreed upon among the collaborators.

At first, 12 participants were run for the O-room, after which the evidence was checked for whether there was a difference between across and within room associations. The sample size was then increased by increments of 4 until either a BF_{10} of 6 or 1/6 was reached, or the maximum sample size of 36 was reached, based on the remaining balance on the Prolific account at that time (£431.66). If evidence was found for a boundary effect for O-rooms, the procedure would be repeated for M-rooms (starting at twelve and incrementing by four) with the same stopping rule (i.e., subject to remaining Prolific balance).

The test determining the termination of the sequential design was a paired *t*-test between within and across room associations for each room-type. This analysis was collapsed across question-type, but restricted to the first block because of the possibility that the second block would not yield an effect due to interference from the first block.

In order to assess whether this data collection plan had enough chance to provide conclusive (moderate) evidence (BF_{10} > 6 | < 1/6), a simplified simulation was ran with directional Bayesian *t*-tests with minimum sample size of 12 and maximum of 36 with effect sizes d = 0, 0.44 and 0.78 that correspond to a null effect, the effect size for O-rooms and the effect size for M-rooms, respectively, that were observed in Experiment 1c (see Figure 4.5).

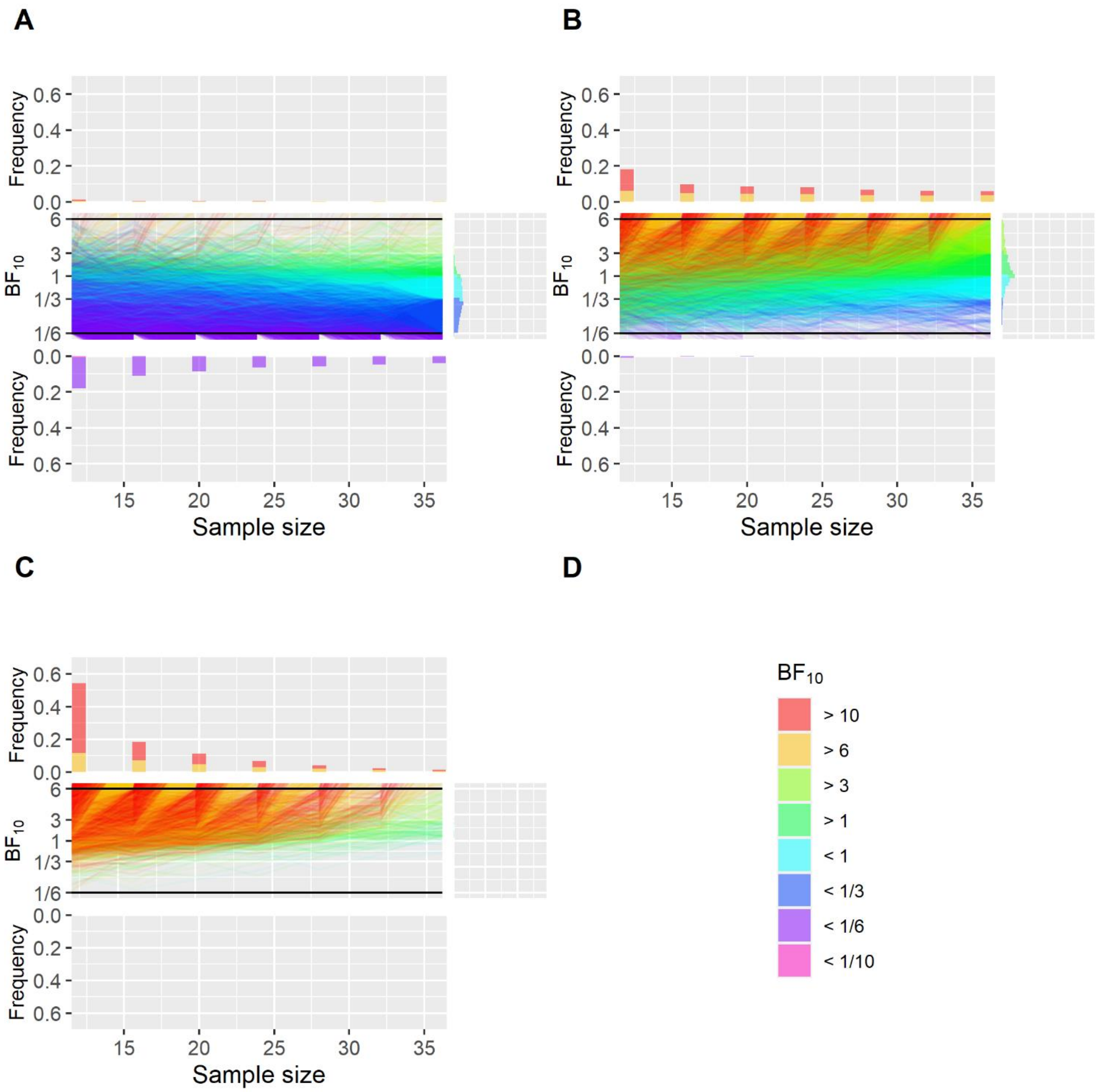

Figure 4.5: Results of design analysis for a Bayesian sequential design: Panel A, B and C show the results of the design analysis for true effect sizes of d = 0.00, d = 0.44 and d = 0.78 respectively. D shows the rainbow colour legend for BF_{10}. Each line in the centre graphic illustrates one simulation of a sequential data collection run. The colouring is determined by the BF_{10} at the end of a simulation when the algorithm terminated either by reaching an evidence criterium (BF_{10} > 6 or BF_{10} < 1/6) or the maximal sample size (N = 36). Histograms at the top and the bottom show the frequency with which simulation end at a particular sample size. Colouring again indicate the strength of the evidence. The right hand side histograms illustrate how many simulation ended by reaching the maximal N.

Table 4.1: Summary of design analysis simulation.

Effect size	Median sample size	Evidence rate	Misleading evidence rate
d = 0.00	28	0.5833	0.0301
d = 0.44	28	0.6321	0.0217
d = 0.78	12	0.9863	0.0002

Thus the probability of exceeding a BF_{01} of 6 when there is no effect, or exceeding a BF_{10} of 6 when the effect size matches that for the O-rooms in Experiment 1c, is around 60% in both cases (with a probability of misleading evidence of 3% or less; see Table 4.1).

Sample

In total, 16 participants (7 female and 9 male), recruited through Prolific, completed this version of the experiment, before the the priori BF criteria was exceeded. Their mean age was M = 27.9 (SD = 10.1) years.

Results

Data collection terminated at N = 16, all of whom completed the O-room condition, at which the BF for the null exceeded the criterion that there is no boundary effect for the first block for accuracy with O-rooms, BF_{01} = 7.06, *d* = -0.252 (see Figure 4.6A). This result also held true when collapsing across both blocks [across: M = 0.398 (SD = 0.0744), within: M = 0.372 (SD = 0.0834), BF_{01} = 7.81, *d* = -0.308]. Overall, memory performance was low, but above chance, BF_{10} = 14.68, M = 0.385 (SD = 0.0667).

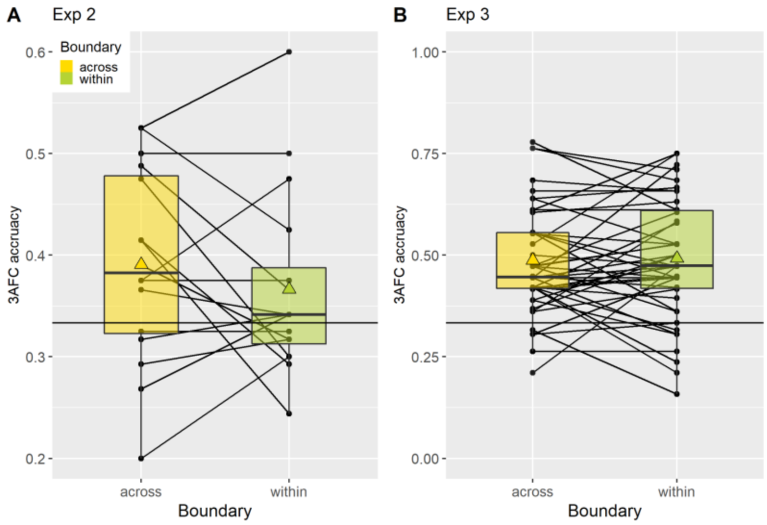

Figure 4.6: Main results for Experiment 2 & 3, where only O-rooms were used: Standard boxplots with triangles illustrating condition means. Note that for panel A, only data from the first block were included. Dots connected by a line belong to the same participant.

The ANOVA with boundary-type, question-type and block did not produce strong evidence for main effects or interactions (all $BF_{01} < 4.5$ and > 0.593), apart for the main effect of question type, $BF_{10} = 9.23$, with performance for the "after" question, M = 0.416 (SD = 0.0785), being higher than for the "before" question, M = 0.355 (SD = 0.0846).

Discussion

While Experiment 1c provided no clear evidence for or against a boundary effect in O-rooms, Experiment 2 found moderate evidence against such an effect. Overall, temporal order memory was above chance, though it was still low. In a final attempt, Experiment 3 tried to increase overall performance. Interestingly, like reported by Horner et al. (2016) performance was better in forward direction (i.e. for the "after" question) showing that are not just non-replicable effects.

4.4 Experiment 3

Experiment 3 also only tested O-rooms. Mean memory performance for these rooms in Experiment 2 was 0. 385 (chance = 0.333), whereas in Horner et al (2016), it was slightly higher, on average 0.44 across experiments. The main aim for Experiment 3 was therefore to further improve overall memory performance.

Three major changes were made to address this: a) memory encoding in Experiment 3 was made intentional rather than incidental (participants were told at the outset that “after watching the video we will also test your memory performance. So, please try to remember the objects that you see and their order”), b) there were two study-test cycles (like in Horner et al., 2016, Experiment 2 & 3), so that only 44 objects at a time were encoded before memory was tested, and c) the rooms were made even more distinct. Regarding the latter case, only five different wall colours and five different floor textures were used in Experiments 1 and 2, which means a room was only unique by the virtue of combining wall colours and floor textures. To increase room distinctiveness in Experiment 3, each floor featured a unique texture that was selected to stand out (e.g. brightly coloured tiles, noticeable carpet patterns, etc.).

The other procedural details including the foil selection were the same as in Experiment 1c. Due to the availability of foils the number of trials per condition (e.g. within vs. across) was either 36 or 38 per cycle. Furthermore, I again decided to only test one question type (the after question) to reduce the number of counter balance conditions and this factor did not interact with any effect in my experiments so far and it also did not interact with the boundary effect in Horner et al. (2016).

Sample

One participant was excluded because they did not complete the whole task. In total, 49 participants (19 female and 30 male), recruited through Prolific, completed this version of the experiment, with mean age of M = 29.2 (SD = 11.8) years. The sample size was determined by the amount of money left for this project. Since the sample size is larger than what was simulated in the design analysis for Experiment 2 (see Table 4.1), power to provide strong evidence was improved relative to that experiment.

Results

Average transformed (overall) accuracy was treated as an outlier if it was above/below two Median Absolute Deviations (i.e. MAD) around the median (see Leys et al., 2013). This led to the exclusion of 3 participants who were above this cut-off. Note this exclusion was not done in the previous experiments because their sample sizes were too small for accurate estimation of spread across participants, but this might also contribute to higher mean performance. For analysis, data from both study-test cycles were collapsed.

Mean accuracy did increase relative to Experiment 2, up to 0.49 (0.26), which was clearly above chance $BF_{10} > 1x10^8$. However, there was again anecdotal evidence against a boundary effect, $BF_{01} = 5.44$, $d = 0.0255$ (see Figure 4.6B).

Discussion

By increasing distinctiveness of the rooms, reducing the memory load via two study-test cycles and instructing participants to remember the objects and their order, Experiment 3 succeeded in improving temporal order memory (such that it was now higher than the average in the Horner et al., 2016, experiments). Nonetheless, there was still evidence against a boundary effect on temporal order memory across O-rooms. There was insufficient time to perform further experiments to narrow down other possible differences with the original study of Horner et al. (2016), but some possible reasons are considered below.

4.5 General discussion

Across five experiments, there was either no evidence for, or evidence against, a boundary effect for O-rooms similar to those used by Horner et al. (2016). In Experiments 1a-1b, where foils for the temporal order judgment were constrained to be close in time to the target, overall performance was not even above chance, which prevents clear interpretation. However, when the foils were randomised (as in the Horner et al., 2016, paper), and overall performance was above chance, there was either no evidence (Experiment 1c) or anecdotal to moderate evidence against (Experiment 2 and 3), a boundary effect with O-rooms.

Surprisingly, there was evidence for a boundary effect in M-rooms in Experiment 1c, but since the effect in M-rooms was predicted to be smaller, if anything, than in O-rooms, this finding was not pursued further.

Potential reasons for the null effect

There are two major differences remaining between the experiments reported in this chapter and the experiments reported in Horner et al. (2016). First, in the Horner et al. (2016) study, navigation was active, whereas here participants merely watched a video of someone else navigating the rooms. Second, the position of the tables and the doors in Horner et al. (2016) varied (such that the rooms formed a closed circle), whereas here the spatial layout of each room was identical (apart from being an O or M-room), which meant that they formed a linear route overall. One possibility is that doors function as a boundary (for temporal order memory at least) only when someone is actively navigating, and possibly trying to remember an overall layout. Another possibility is that the different layouts of rooms used by Horner et al (2016) made them more distinctive.

Following the results of a recent study (Logie & Donaldson, 2021), the idea that the lack of active navigation is the main culprit for the null finding is unlikely, because these authors reported boundary effects for passive movement. In their study, they found that introducing boundaries increased the total number of words learned in a VE, consistent with EST. However, they also found that spatial boundaries were not even necessary, since temporal gaps sufficed to get the memory boost.

In a number of studies on the related location updating effect, it has been shown that when people walk through a door, they forget information (Lawrence & Peterson, 2016; McFadyen et al., 2021; Pettijohn & Radvansky, 2018a, 2018b; Radvansky et al., 2010, 2011; Radvansky & Copeland, 2006; Seel et al., 2019; though note that McFadyen et al., 2021, failed to replicate this effect). With such a paradigm, Pettijohn & Radvansky (2018a) provided direct evidence that boundaries effects are smaller for passive navigation but still detectable. Yet, McFadyen et al. (2021) also found a boundary effect (in terms of a location updating effect), but only for passive movement and when working memory load was high. Furthermore, passive versus active navigation can also be interpreted as different levels of "immersion" in a VE. Immersion was manipulated by Radvansky and colleagues when studying that effect (Radvansky et al., 2011). These researchers typically use a large screen (e.g. Radvansky & Copeland, 2006) that affords a high level of immersion, but the effect also held when a standard screen was used (Radvansky et al., 2011). Another study showed that even only mentally walking through a doorway produced a boundary effect (Lawrence & Peterson, 2016). All of this strongly suggests that the lack of active navigation, while it may reduce the effect, is not the primary reason why the present experiments failed to find a boundary effect.

This leaves the possibility that the experiments described here failed to find an effect because they used a linear environment, where each room followed a single, straight path. In contrast, each study that found boundary effects of any kind (deleterious or beneficial, e.g., for item memory and for temporal order memory respectively) used non-linear environments (Horner et al., 2016; Lawrence & Peterson, 2016; Logie & Donaldson, 2021; Pettijohn & Radvansky, 2018a, 2018b; Radvansky et al., 2010, 2011; Radvansky & Copeland, 2006; Seel et al., 2019). A notable case here is the study from McFadyen and colleagues (2021), where the VE was highly immersive and non-linear, but each hexagonal room was otherwise identical (same layout and colouring etc.), which largely failed to replicate the location updating effect and when they did, it manifested in increased false alarms rate instead of the expected reduction in hit rates. All in all, it seems that having a complex environment with distinct rooms is a necessary condition to find robust boundary effects. Complex environments and distinct rooms might entail difference orientation responses (e.g. where is the next door) when entering a room and might also be more conducive to forming rooms schemas, which might affect the size of the boundary effect.

Another less interesting, procedural difference was the present use of online testing rather than in-person testing of Horner et al (2016). However, this seems unlikely to be important, given that memory performance was comparable to Horner et al (2016), and indeed better, in the present Experiment 3.

Despite the fact that the usefulness of item-item associations in sequential reproduction has been questioned (see Henson et al., 1996), when DuBrow & Davachi (2013, Experiment 2) de-emphasised associative processing between successive items, they found that they were unable to reproduce their boundary effect on temporal order memory (based on a 2AFC recency judgement). While a lack of item-item associations might account for the inability to find a boundary effect in Experiments 1 & 2, it seems less likely for Experiment 3, where participants where explicitly instructed to remember the objects they see and their order.

In summary, the original aim elucidating which factors (PC or PE) contribute to the effect of spatial boundaries failed because I was unable to replicate the original finding on which this paradigm was based and which was supposed to serve a control condition. I therefore never got to use M-rooms to manipulate different types of predictions and their effects on temporal order memory.

More generally, recent work suggests there might not be anything special about doors (Logie & Donaldson, 2021; McFadyen et al., 2021). If walking through a doorway is sufficient to produce boundaries effects, then these should have also manifested in my experiments; however it seems that this is not the case, so other factors must contribute. It is important to keep in mind that there was always a sizeable PC when walking into the next room, and also at least some degree of PE[11] in all experiments discussed in this chapter. Despite the presence of PC, as well as at least some PE, there was no boundary effect for O-rooms. It could therefore be that a boundary effect only arises when PE is stronger as it is possibly the case when the room layout changes, which naturally requires different orientation responses compared to when the table is always in the same space.

[11] As participants did know what the next room would look like, there was always some PE. In Experiments 1 and 2, the number of wall colours and floor textures was only five, so participants might have formed corresponding expectations, but in Experiment 3 each rooms was unique, in terms of the floor textures and wall colours.

5 DISCUSSION

PIMMS is a powerful framework to explain how PE drives memory encoding. It describes the interaction of perceptual, semantic and episodic memory systems and was originally crafted to explain how recollection and familiarity arise based on a predictive perspective on the brain. Its focus on the role of PE in memory encoding makes the model perfect to be applied to the literature of novelty and surprise, which is extensive but somewhat incoherent (Quent et al., 2021). When doing so, it helps to debunk the idea that novelty *per se* is associated with memory improvement, and provides strong grounds to distinguish the two concepts of novelty and surprise.

Despite the success of PIMMS, there are several phenomena surrounding novelty, PE and memory encoding that are not explained by PIMMS, even though they are of high theoretical relevance. In my thesis, I addressed three such phenomena: 1) how memory encoding can be improved when information is congruent with prior knowledge (i.e., with little PE), forming a U-shaped function of expectancy; 2) how PIMMS is silent on neurobiology and therefore cannot explain phenomena like behavioural tagging that are believed to improve memory; and 3) how PIMMS - though it focusses on PE - is insufficient to explain how our continuous experience is segmented into episodes by event boundaries, which have important consequences for memory performance.

5.1 U-shape between expectancy and object-location memory

As predicted by SLIMM – a more specific neuroanatomical model developed partly in response to the short-comings of PIMMS – I found that there is indeed a U-shaped relationship between memory and expectancy, at least in the case of memory for locations of objects in a virtual kitchen as a function of the expectancy of those locations. This U-shape emerged in the recall task, as well as to some extent in the 3AFC recognition task. Even though the evidence from the recognition task was weaker, and did not hold for the interrupted regression analysis, the consistent quadratic component in the recognition task is important because the foils controlled for the guessing bias that was found when participants recalled objects at expected locations. Furthermore, recollection in the recognition task (as estimated for example by the proportion of remember judgments) clearly showed a U-shape in all analyses, including the interrupted regression.

The U-shape for recollection was not only unpredicted according to the PIMMS framework, but also according to the more specific SLIMM model. SLIMM predicts that recollection and familiarity would play different roles for memory performance at both ends of the continuum, with recollection associated with memory for unexpected information, but familiarity associated with memory for expected information. While one possibility is that SLIMM is wrong, and memory for expected information is also associated with recollection, another possibility is that the spatial/associative nature of the memory probed in the experiments of Chapter 2 invariably engages recollective mechanisms, since recollection and spatial memory are closely related, and both rely on the hippocampus (see e.g. Burgess et al., 2002).

However, other studies have shown that memory for non-spatial information can be associated with recollection even when congruent with expectations. For instance, Amer et al. (2018, 2019) presented participants with realistic and unrealistic prices of objects, and found that those prices that were congruent with the participants' prior knowledge were associated with high confidence memory responses, which suggests that recollection might be involved (though see discussion in Chapter 2 about the relationship between confidence and recollection, since it is also possible that schema-congruent items are recognised with high confidence without recollection).

Another issue that is likely to be relevant is the nature of the information recollected. Recollection normally refers to incidental contextual information, such as remembering when/where information was encountered, or what one happened to be thinking about at that time. It may be useful to distinguish between information that is schema-relevant versus schema-irrelevant (which may relate to the distinction between "intrinsic" and "extrinsic" context, Godden & Baddeley, 1980). Recent evidence from a VR version of this paradigm supports this claim by showing that the context-dependent memory effect was only observed for words that were related to an active schema (Shin et al., 2021). According to SLIMM, it is schema-irrelevant information that is *not* encoded for expected events, i.e., schema-relevant information is still encoded (while for unexpected events, all such concurrent information is encoded in an "episodic snapshot"; van Kesteren et al, 2012). Thus if participants base their "remember" response on schema-relevant information, then such responses would be predicted to increase for expected as well as unexpected information.

The question then becomes whether the location of objects in the virtual kitchen in the experiments of Chapter 2, or the price of objects in the Amer et al. (2018, 2019) studies, are examples of schema-relevant information. For example, one could argue that the typical location of a kettle (in a kitchen, on a sideboard) is part of our schema for kitchens and kettles, in which case SLIMM predicts that location would be retrieved, and hence could be used to justify a “remember” response. Indeed, participants were told that memory for object location would be tested later. By contrast, the (relative) time at which a kettle was encountered while exploring a kitchen, or perhaps the colour of the kettle, may be irrelevant to the schema, in which case SLIMM predicts they would not to be encoded, provided the kettle appears where expected (whereas such details would be encoded as part of the episodic snapshot if the kettle were at a surprising location). Likewise, the typical price of objects could be considered as intrinsic to the knowledge (schema) for those objects, and hence explain why they were recognised with high confidence (and possibly frequent recollection) in Amer et al.’s studies. If this is true, then by re-defining (to participants) the nature of retrieved information that should be used for a remember response, or by testing source memory for incidental perceptual details like the colour of an object (if irrelevant to the schema for that object), might reveal a different pattern, with only an increase towards unexpected locations (no U-shape). This could be tested in future experiments.

Having said this, there are other studies that suggest that the improved memory for expected information is accompanied by recollection even for incidental (schema-irrelevant) details. Using congruent and incongruent (or unrelated) noun-adjective pairings for example, such as “*yellow banana*” versus “*yellow strawberry*” (Bein et al., 2015; Reggev et al., 2018), Bein and colleagues (2015) reported a congruency advantage that was associated with increased number of remember but not of know judgements for target words (e.g. banana). Explaining this finding requires understanding exactly what type of information participants used as the basis of their remember judgments: if it were incidental information, then this would indeed be inconsistent with SLIMM.

Evidence for the U-shape was weakest for accuracy in the 3AFC recognition task, so future work should be conducted to provide stronger evidence in this regard. A simple way to do this could be to present videos to participants, instead of presenting the environment in iVR, and collect data online for the recognition task only. This is easily achieved with the VR technology because the exact position and orientation of the head-mounted display (HMD) can be tracked in the physical world and saved. The tracked positions of an HMD can be then used to create videos where a real person explores the room freely and in a realistic fashion. This ensures that a) every participant experiences the room in the same way visually but also b) that objects and locations can be easily shuffled so that many more sets can be tested than would be practical with in-person iVR testing. In fact, this would also address another problem, which is because participants explored the kitchen freely, they sometimes missed objects because they simply did not explore specific parts of the room. Generalising the results across a larger set of object-locations is especially important given the suggestion in Chapter 2 that the results might depend on stimulus set. Relatedly, it would also of course be interesting to generalise the results to other room types (e.g. a bathroom or a living room).

The observed U-shape should also be studied as a function of various other factors like retention interval (e.g. one could test memory the next day, to see if consolidation processes moderate the effect), or participant age (e.g., if the mPFC vs MTL systems of SLIMM develop in children, or deteriorate in old age, at differential rates). For instance, it might be expected that the relative advantage for expected versus unexpected information might increase after a delay, as memory becomes more gist-like (or episodic details are forgotten), which has been repeatedly claimed (Sekeres et al., 2018; Winocur & Moscovitch, 2011). Similarly, slower maturation in children, or faster deterioration in old age, of the MTL system might similarly attenuate the advantage for unexpected information (e.g. Raz, 2005). Furthermore, increasing world knowledge as people get older might actually accentuate the U-shape, i.e., result in greater advantages for both expected and unexpected information, as the predictions from schema get stronger.

Other potential uses of the iVR paradigm: effects of stress

Another interesting future avenue could be to investigate how stress (either induced or via direct administration of stress hormones) affects the U-shaped relationship of expectancy. An important prediction of SLIMM is that different brain regions (MTL and mPFC) play differential roles in supporting the schema-incongruent vs schema-congruent advantage for memory. Manipulating stress levels might help to provide more direct evidence for the role of the hippocampus in the incongruent advantage.

As a brief background, there are two receptor types that bind the stress hormones cortisol (humans) and corticosterone (rodents): mineralocorticoids (MR; Type I) and glucocorticoids (GR; Type II; Reul & de Kloet, 1985). Both receptors types can be found in the hippocampus. So unsurprisingly, there is consistent evidence that the hippocampus is strongly affected by stress (Patel et al., 2000; Sánchez et al., 2000). Studies of the temporal dynamics suggest that the hippocampus' reaction to stress first comprises excitation and then inhibition (Diamond et al., 2007). This pattern was confirmed by an fMRI study that examined the time-course of the BOLD response after hydrocortisone injection: hippocampal variance first increased (peaking at 5-10 min) and then decreased, with a peak 30-35 min post injection (Lovallo et al., 2010), which offers intriguing time-windows during which the iVR paradigm could be completed.

With the close relationship of stress and hippocampus in mind, it is also not surprising that an abundance of research has shown that stress is a potent modulator of (episodic) memory (Schwabe et al., 2012; Shields et al., 2017; Wolf, 2008, 2009). On the behavioural level, whether stress enhances or impairs memory depends on the timing relative to the stage of memory formation according to the "integrated" model (Schwabe et al., 2012; Vogel & Schwabe, 2016): stress impairs if experienced long before encoding, but enhances if only experienced shortly before it. On the other hand, stress before retrieval has a detrimental effect. These predictions were confirmed by a meta-analysis (Shields et al., 2017).

To my knowledge, only two studies specifically looked at the effect of stress on schema-related memory. The first study showed that stress and hydrocortisone can disrupt the benefit of prior knowledge (Kluen et al., 2017). In this study, participants learned the ordinal relationship of a number of galaxies on the first day and then learned two new hierarchies on the second day. One half of one hierarchy consisted of galaxies whose ordinal relationship were learned on Day 1. Hence if participants used the knowledge they acquired, it should give them a performance advantage. Psychosocial stress immediately and 25 minutes before the task on the second day, as well as hydrocortisone 45 minutes before the task, impaired participants' ability to use that knowledge. Interestingly, explicit hierarchy recall was not impaired in participants who received hydrocortisone, which might suggest that the results are not due to a simple retrieval deficit.

In the second study (Vogel et al., 2018), participants underwent a psychosocial stress test 45 minutes before encoding schema-related and schema-unrelated words in a scanner. Three to five days later, participants' memory was tested. While stress did not affect memory directly, participants showed better memory for schema-related words. While participants in the non-stressed control group showed increased activation in the mPFC and reduced activation in the hippocampus, participants in the stress group showed increased hippocampal activity during schema-related processing. This increased hippocampal activity was positively related to the magnitude of the cortisol response and interestingly to the number of schema-related false alarms

Both studies did not directly address the question of whether stress impacts processing of incongruent information more than the processing of congruent information. This requires at least three ordered schema-related conditions (incongruent, neutral and congruent) or a more continuous paradigm like that in Chapter 2. To demonstrate that the hippocampus is especially important for the incongruency benefit, one could administer hydrocortisone to participants / subject them to stress before they complete the iVR experiment. The timing should be arranged so that the iVR task is completed during the refractory period (see Lovallo, Robinson, Glahn, & Fox, 2010), during which hippocampal learning is impaired, which in turn should lead to a reduced advantage for objects at unexpected locations. This is just one example of how the iVR paradigm I developed in Chapter 2 has potential for further research.

5.2 Experiencing iVR for the first time does not improve memory

As discussed earlier, besides not predicting a memory advantage for expected information, PIMMS is also silent on neurobiology. Due to this silence, PIMMS cannot address important aspects of how novelty and surprise relate to memory consolidation. One such important phenomenon is behavioural tagging, which explains how experiencing something novel (or significant) can boost memory for events experienced in close temporal proximity, due to neurochemical changes at synapses that happen around that time. The idea of behavioural tagging developed from the general idea of tag-and-capture (Frey & Morris, 1997; Redondo & Morris, 2011). Evidence for this phenomenon comes from non-humans (Ballarini et al., 2009; Moncada & Viola, 2007) and humans (Ballarini et al., 2013; Fenker et al., 2008; Ramirez Butavand et al., 2020).

In Chapter 3, I predicted that the novelty of one’s first experience of iVR (while performing the same tasks as in Chapter 2) would enhance memory for preceding, unrelated information. Contrary to my expectations, I found evidence in favour of the null hypothesis (i.e. iVR did not improve memory for preceding words). While I did not necessarily expect an effect on immediate recall, given that behavioural tagging presumes that a period of consolidation is needed (see for instance Roig et al., 2016), there was evidence against any effect for delayed recognition as well as immediate recall. This was at odds with previous VR studies, as reprised below. The main aims of this study were to examine whether novelty would differentially affect weakly or deeply encoded information, and whether it would affect recollection and familiarity in different ways. The corresponding BFs for these hypotheses were not conclusive. At least two possible reasons for the lack of evidence for any effects of novelty (as discussed in Chapter 3) were that: a) the novelty manipulation might not have been strong enough (though this seems unlikely given that other reports found this effect with merely presenting unseen images); and b) the Control and the Novelty group differed not only in terms of experiencing something the first or the second time, but also in terms of the difficulty of the task that needed to be completed during that experience. Completing a difficult task after learning has been shown to negatively affect memory performance (Dewar et al., 2007; Wixted, 2004). While counting the objects within the virtual kitchen may not be a difficult task, the recall and recognition tasks were reasonably demanding, so could have disrupted consolidation and reduced any effects of novelty on that consolidation. Nonetheless, it is interesting to note that novelty is often associated with increased cognitive demands / exploration, suggesting that it would be important for future experiments to factorially cross novelty with task difficulty (see Chapter 3).

In general, evidence for behavioural tagging in humans is still scarce, with several other recent null findings (Biel et al., 2020; Biel & Bunzeck, 2019). Nonetheless, there is also recent study that did find evidence for behavioural tagging in high-school students (e.g. Ramirez Butavand et al., 2020), plus a further study that used familiar and novel Minecraft environments and found a retroactive effect, but only for ADHD patients and not for typically developing children/adolescents (Baumann et al., 2020). Another study failed to find an effect of surprising actions within video clips on memory for other actions that happened before (Ben-Yakov et al., 2020). While the evidence in animal studies is more consistent, it is worth noting that the proactive effect of VR on human memory reported by Schomaker and colleagues (2014, 2021) is unlikely to represent the behavioural tagging-like processes seen in animal experiments, because the latter is assumed to take time to influence memory consolidation, whereas Schomaker and colleagues tested memory immediately after the VR experience.

Despite the heterogeneous state of the literature on retroactive/proactive effects of novelty on memory, the idea of behavioural tagging remains strong, and extends to manipulations other than novelty, such as manipulations of post-encoding stress (Lopes da Cunha et al., 2018; Quent et al., 2018; Ritchey et al., 2017), fear conditioning (Dunsmoor et al., 2015; Hennings et al., 2021), physical exercise (Roig et al., 2013, 2016), post-encoding arousal (e.g. Nielson & Arentsen, 2012) and reward (e.g. Patil et al., 2017). While there are several indirect demonstrations that (behavioural) tagging affects weakly-encoded information (Baumann et al., 2020; Lopes da Cunha et al., 2018; Quent et al., 2018)[12], further systematic work is needed to provide direct evidence for the secondary hypotheses tested in Chapter 3, i.e., that tagging should benefit weak memories more than strong ones, and differentially affect the subsequent experience of recollection versus familiarity.

In summary, as noted above, one fruitful avenue for future work would be to manipulate novelty in ways that are not confounded by cognitive demand (see discussion in Chapter 3). However, the behavioural-tagging hypotheses can also be tested by using other manipulations like stress, arousal, physical exercise, that may be easier to de-confound from cognitive demand, and combined with the weak/strong encoding task that I used in Chapter 3. It is also important to pre-register and report such future experiments (e.g. as my Registered Report version of Chapter 3), just in case the positive effects in the literature are false positives, and many other negative results are simply not reported (and to get a more accurate indication of the size of any effect, e.g., for meta-analyses).

5.3 No effect of (room) boundaries on temporal order memory

In Chapter 4, I described a paradigm that was originally designed to study what aspects of walking through a doorway (perceptual change, PC, and/or prediction error, PE) leads to previously reported effects of boundaries (like doorways) effects on memory. Unfortunately, this failed because I was unable to replicate the findings of a very similar study (Horner et al., 2016). Chapter 4 discusses several reasons that might have led to this failure to replicate. One likely culprit is that the environment that I created was merely a chain of rooms that were arranged on a straight line, while Horner et al. used a more complex arrangement of rooms (i.e., more complex layout of the whole environment).

[12] Note that the work from Lopes da Cunha et al. (2018) and from Baumann et al. (2020) involved groups of people for which weaker encoding can be assumed (ADHD patients or student group with lower memory in matched control group) but was not manipulated, hence this was classified as indirect evidence.

While a large number of studies have shown boundary effects (not only on temporal order memory, but also other types of memory) after walking through a doorway (Horner et al., 2016; Lawrence & Peterson, 2016; Logie & Donaldson, 2021; Pettijohn & Radvansky, 2018a, 2018b; Radvansky et al., 2010, 2011; Radvansky & Copeland, 2006; Seel et al., 2019), there are also some recent suggestions that doors might not be special after all. For instance, Logie and Donaldson (2021) reported that temporal gaps showed a comparable boundary effect to walking through doors. Furthermore, a study on the location updating effect revealed that simply using identical rooms - instead of rooms that varied for example with respect to wall colours, as it is typically done - caused the boundary effect on item memory to vanish (McFadyen et al., 2021). It is conceivable rooms that are more unpredictable in terms of the layout and the interior design demand stronger orienting responses. Imagine you enter a room where you first have look for the exit versus a room that has the exact same layout and/or design as the ten previous rooms. It is possible that boundary effects only arise in the scenario where there is some level of significant unpredictability. When taken together with my results, these data suggest that a door per se is not the cause of boundary effects, and it is more likely that other factors like PC and/or PE drive these effects. Ironically, this is exactly what was planned for my series of experiments, but which failed because I could not find a basic effect in the standard condition (i.e. walking through the doorway).

For future work, even though passive navigation is most likely *not* the main culprit for the null finding, the ease with which the experiment from Chapter 4 can be changed from passive to active navigation with the available code potentially warrants another attempt. While most studies still found a boundary effect with passive navigation, some also report a reduction of the effect (though see McFadyen et al., 2021). Thus counteracting this reduction might be sufficient to find a boundary effect for the paradigm in Chapter 4.

Another possibility is to change the layout of the rooms. Originally, the O and M-rooms in Chapter 4 were created because they can be easily be placed one after the other on a straight line, as this made scaling the environment (i.e. adding more rooms) easier. Furthermore, M-rooms were designed so that if a person is anywhere within one half of the room, they cannot see into the other half. This was done to ensure that I could change the wall colour in the other half of the room without that being visible before the participant entered that part of the room. Nevertheless, it is possible to move the doors (and possibly the tables) and create a more complex environment that comes with a certain level of unpredictability concerning the layout of the encountered rooms. By simply changing the locations of the doors and tables within both room types, it is possible to construct a series of rooms that is not simply a linear track. This might be more conducive to finding the boundary effects that Horner et al. (2016) reported.

Even if doorways do not always serve as boundaries, the effect of boundaries on cognition and the brain in general are manifold (for recent reviews see Clewett et al., 2019; Richmond & Zacks, 2017). A myriad of studies also highlighted that other types of boundaries affect memory (DuBrow & Davachi, 2013, 2016; Ezzyat & Davachi, 2014; Heusser et al., 2018; Rouhani et al., 2020; Swallow et al., 2009). Despite the replication failure in Chapter 4, these effects are not fully explained by PIMMS: while PEs may cause event boundaries (Zacks et al., 2007), how that PE affects memory for surrounding information (e.g., the relative order of items that straddle versus do not straddle that PE) is not currently specified by PIMMS. It may be fruitful therefore to combine PIMMS with other event-based frameworks like EST (Reynolds et al., 2007; Zacks et al., 2007), EHM (Radvansky et al., 2011) - see Chapter 1 – as well as the more general Temporal Context Model (TCM; Norman et al., 2008; Polyn et al., 2009). Indeed, it is important for any general theory of memory to be able to incorporate working memory, as well as long-term memory, as well as explain effects that are commonly attributed to a dynamic temporal context (e.g. see Clewett et al., 2019), including the role of consolidation (Yonelinas, 2019), as elaborated below.

5.4 Further gaps in PIMMS

Beyond the limitations that I addressed in Chapter 2 to 4, there are several other gaps that a comprehensive framework of (declarative) memory needs to address. I finish with mentioning a few of these, that are most relevant to the work in this thesis. The first of these, related to Chapter 2, is that, while PIMMS does consider how prior (semantic) knowledge interacts with the perceptual and with episodic memory system, it does not describe how semantic knowledge is acquired in the first place. Even if there is no inherent novelty benefit (see Quent et al., 2021 or Chapter 1), we are still able to learn truly novel information. For this, McClelland and colleagues suggested the “complementary learning systems” model (Kumaran et al., 2016; McClelland et al., 1995), which provides an algorithmic model how new information is slowly acquired by the cortex. In that framework, acquisition of truly novel knowledge is accomplished with help of a fast learner that is based in the hippocampus (similar to the episodic memory system) and a slow learner that is based in the neocortex (similar to the semantic memory system). New semantic knowledge has to be acquired by prolonged interleaved learning, which abstracts statistical regularities from individual episodes, for which hippocampal replay is supposed to play a crucial role. This kind of mechanism is lacking in PIMMS.

PIMMS is also simplistic in terms of the brain regions that are considered. A comprehensive model would not only include the mPFC, as suggested by SLIMM, but several other regions. The locus coeruleus (LC), the substantia nigra, striatum and ventral tegmental area (SN/VTA) all appear to be important for novelty and memory formation (Clos et al., 2019; Duszkiewicz et al., 2019; Fenker et al., 2008; Guitart-Masip et al., 2010; Hollerman & Schultz, 1998; Kamiński et al., 2018; Mikell et al., 2014; Murty et al., 2016; Schott et al., 2004; Wittmann et al., 2007). In relation to Chapters 2 and 3, one study showed, for example, that context surprise can induce early phasic responses in the VTA, while a second peak in the VTA could reflect a feedback signal to the hippocampus (Mikell et al., 2014). It has been further proposed that a functional loop exists between these midbrain regions and the hippocampus, via the ventral striatum (nucleus accumbens; Kamiński et al., 2018; Lisman & Grace, 2005 see below). These kind of interactions are neglected by PIMMS.

Other studies have highlighted further brain regions, including the bilateral ventral striatum and bilateral putamen/insula, that track signed, reward PE (Pine et al., 2018), and the importance of areas like the SN/VTA for reward PE has been known for over two decades (Hollerman & Schultz, 1998; Schultz & Dickinson, 2000). It seems like the same populations of dopamine neurons that respond to reward PE may also be sensitive to context surprise (in the absence of an explicit reward; see Hollerman & Schulz, 1998). Reward anticipation (e.g. Wittmann et al., 2007) and reward PE (Ergo et al., 2020) are also known to affect declarative memory, yet PIMMS completely ignores reward as a dimension. Interestingly, the same set of regions is also believed to play a crucial role in event segmentation and error detection (Zacks et al., 2007). It would therefore be interesting to see how PIMMS can be extended when considering these phenomena and brain regions.

Lastly, it has to be acknowledged that PIMMS is only a framework. However, some of its theoretical insights can inform a general theory of memory that not only addresses how the perceptual, semantic and episodic memory systems interact, but also how they relate to working memory. Such a theory also needs to understand perception and memory as ongoing processes, and not simply as snapshots. Finally, formal computational implementations, informed by neurobiology, are needed to make testable, quantitative predictions.

5.5 Conclusion

The interactions between novelty, PE and memory encoding are complex. There is no doubt that prior knowledge has a profound influence on how we perceive the world and how we remember new events. However, to capture the details of this complexity, PIMMS needs to be developed further, to cover, for example, how new schemas are learned (to make predictions), how novelty and surprise can also affect neurobiological processes like consolidation, and how continuous stimuli are segmented for encoding into memory. Nonetheless, PIMMS has proved a helpful framework for thinking about some of these concepts, and designing experiments to demonstrate its short-comings, which is how science proceeds.

6 References

Åberg, C. S., & Nilsson, L. (2001). Facilitation of source discrimination in the novelty effect. *Scandinavian Journal of Psychology*, *42*(4), 349–357. https://doi.org/10.1111/1467-9450.00246

Åberg, C. S., & Nilsson, L. G. (2003). A strict response criterion yields a mirror effect in the novelty paradigm. *Scandinavian Journal of Psychology*, *44*(5), 425–432. https://doi.org/10.1046/j.1467-9450.2003.00363.x

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. *Behavioral and Brain Sciences*, *22*(3), 425–444. https://doi.org/DOI: 10.1017/S0140525X99002034

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203

Amer, T., Giovanello, K. S., Grady, C. L., & Hasher, L. (2018). Age differences in memory for meaningful and arbitrary associations: A memory retrieval account. *Psychology and Aging*, *33*(1), 74–81. https://doi.org/10.1037/pag0000220

Amer, T., Giovanello, K. S., Nichol, D. R., Hasher, L., & Grady, C. L. (2019). Neural Correlates of Enhanced Memory for Meaningful Associations with Age. *Cerebral Cortex*, *29*(11), 4568–4579. https://doi.org/10.1093/cercor/bhy334

Anderson, J. R. (1981). Effects of prior knowledge on memory for new information. *Memory & Cognition*, *9*(3), 237–246.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063–1087. https://doi.org/10.1037/0278-7393.20.5.1063

Ballarini, F., Martínez, M. C., Díaz Perez, M., Moncada, D., & Viola, H. (2013). Memory in Elementary School Children Is Improved by an Unrelated Novel Experience. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0066875

Ballarini, F., Moncada, D., Martinez, M. C., Alen, N., & Viola, H. (2009). Behavioral tagging is a general mechanism of long-term memory formation. *Proceedings of the National Academy of Sciences*, *106*(34), 14599–14604. https://doi.org/10.1073/pnas.0907078106

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.

Bastin, C., Van Der Linden, M., Schnakers, C., Montaldi, D., & Mayes, A. R. (2010). The contribution of familiarity to within- and between-domain associative recognition memory: Use of a modified remember/know procedure. *European Journal of Cognitive Psychology*, *22*(6), 922–943. https://doi.org/10.1080/09541440903007834

Baumann, V., Birnbaum, T., Breitling-Ziegler, C., Tegelbeckers, J., Dambacher, J., Edelmann, E., Bergado-Acosta, J. R., Flechtner, H.-H., & Krauel, K. (2020). Exploration of a novel virtual environment improves memory consolidation in ADHD. *Scientific Reports*, *10*(1), 21453. https://doi.org/10.1038/s41598-020-78222-4

Bayen, U. J., Nakamura, G. V, Dupuis, S. E., & Yang, C.-L. (2000). The use of schematic knowledge about sources in source-monitoring. *Memory & Cognition*, *28*(3), 480–500. https://doi.org/10.3758/BF03198562

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., & Maril, A. (2015). Delineating the Effect of Semantic Congruency on Episodic Memory: The Role of Integration and Relatedness. *PLOS ONE*, *10*(2), e0115624. https://doi.org/10.1371/journal.pone.0115624

Ben-Yakov, A., & Henson, R. N. (2018). The hippocampal film editor: Sensitivity and specificity to event boundaries in continuous experience. *Journal of Neuroscience*, *38*(47), 10057–10068. https://doi.org/10.1523/JNEUROSCI.0524-18.2018

Ben-Yakov, A., Smith, V., & Henson, R. N. (2020). The temporal dominion of surprise. *Psychonomic Bulletin & Review*.

Bergado, J. A., Lucas, M., & Richter-Levin, G. (2011). Emotional tagging—A simple hypothesis in a complex reality. *Progress in Neurobiology*, *94*(1), 64–76. https://doi.org/10.1016/j.pneurobio.2011.03.004

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity.* McGraw-Hill Book Company.

https://doi.org/10.1037/11164-000

Biel, D., & Bunzeck, N. (2019). Novelty Before or After Word Learning Does Not Affect Subsequent Memory Performance. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01379

Biel, D., Steiger, T. K., Volkmann, T., Jochems, N., & Bunzeck, N. (2020). The gains of a 4-week cognitive training are not modulated by novelty. *Human Brain Mapping*, *41*(10), 2596–2610. https://doi.org/10.1002/hbm.24965

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*(2), 177–220. https://doi.org/10.1016/0010-0285(79)90009-4

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717–726.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*(2), 207–230.

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*, 22–31.

Brod, G., & Shing, Y. L. (2019). A boon and a bane: Comparing the effects of prior knowledge on memory across the lifespan. *Developmental Psychology*, *55*(6), 1326.

Bruce, D., & Gaines, M. T. (1976). Tests of an organizational hypothesis of isolation effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 59–72. https://doi.org/https://doi.org/10.1016/S0022-5371(76)90007-4

Bunzeck, N., & Düzel, E. (2006). Absolute Coding of Stimulus Novelty in the Human Substantia Nigra/VTA. *Neuron*, *51*(3), 369–379. https://doi.org/10.1016/j.neuron.2006.06.021

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The Human Hippocampus and Spatial and Episodic Memory. *Neuron*, *35*(4), 625–641. https://doi.org/10.1016/S0896-6273(02)00830-9

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Clewett, D., DuBrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. *Hippocampus*, *29*(3), 162–183. https://doi.org/10.1002/hipo.23074

Clos, M., Bunzeck, N., & Sommer, T. (2019). Dopamine Enhances Item Novelty Detection via Hippocampal and Associative Recall via Left Lateral Prefrontal Cortex Mechanisms. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *39*(40), 7920–7933. https://doi.org/10.1523/JNEUROSCI.0495-19.2019

Cockcroft, J., Berens, S., Gaskell, M. G., & Horner, A. J. (2021). *Schematic information influences memory and generalisation behaviour for schema-relevant and-irrelevant information*.

Cooper, E., Greve, A., & Henson, R. N. (2017). Assumptions behind scoring source versus item memory: Effects of age, hippocampal lesions and mild memory problems. *Cortex*, *91*, 297–315. https://doi.org/10.1016/j.cortex.2017.01.001

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dewar, M. T., Cowan, N., & Sala, S. Della. (2007). Forgetting Due to Retroactive Interference: A Fusion of Müller and Pilzecker's (1900) Early Insights into Everyday Forgetting and Recent Research on Anterograde Amnesia. *Cortex*, *43*(5), 616–634. https://doi.org/https://doi.org/10.1016/S0010-9452(08)70492-1

Diamond, D. M., Campbell, A. M., Park, C. R., Halonen, J., & Zoladz, P. R. (2007). The Temporal Dynamics Model of Emotional Memory Processing: A Synthesis on the Neurobiological Basis of Stress-Induced Amnesia, Flashbulb and Traumatic Memories, and the Yerkes-Dodson Law. *Neural Plasticity*, *2007*, 1–33. https://doi.org/10.1155/2007/60803

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Dobbins, I. G., Kroll, N. E. A., Yonelinas, A. P., & Liu, Q. (1998). Distinctiveness in Recognition and Free Recall: The Role of Recollection in the Rejection of the Familiar. *Journal of Memory and Language*, *38*(4), 381–400. https://doi.org/10.1006/jmla.1997.2554

Draschkow, D., & Võ, M. L. H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-16739-x

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277–1286. https://doi.org/10.1037/a0034024

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory*, *134*, 107–114. https://doi.org/10.1016/j.nlm.2016.07.011

Dudai, Y. (2004). The Neurobiology of Consolidations, Or, How Stable is the Engram? *Annual Review of Psychology*, *55*(1), 51–86. https://doi.org/10.1146/annurev.psych.55.090902.142050

Dunlosky, J., Hunt, R., & Clark, E. (2000). Is Perceptual Salience Needed in Explanations of the Isolation Effect? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*, 649–657. https://doi.org/10.1037/0278-7393.26.3.649

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature*, *520*(7547), 345–348. https://doi.org/10.1038/nature14106

Duszkiewicz, A. J., McNamara, C. G., Takeuchi, T., & Genzel, L. (2019). Novelty and Dopaminergic Modulation of Memory Persistence: A Tale of Two Systems. *Trends in Neurosciences*, *42*(2), 102–114. https://doi.org/10.1016/j.tins.2018.10.002

Ergo, K., De Loof, E., & Verguts, T. (2020). Reward Prediction Error and Declarative Memory. *Trends in Cognitive Sciences*, *24*(5), 388–397. https://doi.org/https://doi.org/10.1016/j.tics.2020.02.009

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, *81*(5), 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042

Fenker, D. B., Frey, J. U., Schuetze, H., Heipertz, D., Heinze, H.-J., & Düzel, E. (2008). Novel scenes improve recollection and recall of words. *Journal of Cognitive Neuroscience*, *20*(7), 1250–1265. https://doi.org/10.1162/jocn.2008.20086

Fernández, G., & Morris, R. G. M. (2018). Memory, Novelty and Prior Knowledge. *Trends in Neurosciences*, *41*(10), 654–659. https://doi.org/10.1016/j.tins.2018.08.006

Foster, M. I., & Keane, M. T. (2015). Why some surprises are more surprising than others: Surprise as a metacognitive sense of explanatory difficulty. *Cognitive Psychology*, *81*, 74–116. https://doi.org/10.1016/j.cogpsych.2015.08.004

Foster, M. I., & Keane, M. T. (2019). The Role of Surprise in Learning: Different Surprising Outcomes Affect Memorability Differentially. *Topics in Cognitive Science*, *11*(1), 75–87. https://doi.org/10.1111/tops.12392

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nat Rev Neurosci*, *6*(2), 119–130. http://dx.doi.org/10.1038/nrn1607

Frey, U., & Morris, R. G. M. (1997). Synaptic tagging and long-term potentiation. *Nature*, *385*(6616), 533–536. https://doi.org/10.1038/385533a0

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, *18*(1), 23–30. https://doi.org/10.3758/BF03202642

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory*, *10*(2), 83–98. https://doi.org/10.1080/09658210143000281

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4). https://doi.org/10.1214/08-AOAS191

Geraci, L., & Manzano, I. (2010). Distinctive items are salient during encoding: Delayed judgements of learning predict the isolation effect. *Quarterly Journal of Experimental Psychology*, *63*(1), 50–64. https://doi.org/10.1080/17470210902790161

Godden, D., & Baddeley, A. (1980). When does context influence recognition memory? *British Journal of Psychology*, *71*(1), 99–104. https://doi.org/https://doi.org/10.1111/j.2044-8295.1980.tb02735.x

Greene, R. L. (1999). The role of familiarity in recognition. *Psychonomic Bulletin & Review*, *6*(2), 309–312. https://doi.org/10.3758/BF03212335

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001

Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, *148*(2), 325–341. https://doi.org/10.1037/xge0000498

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97. https://doi.org/https://doi.org/10.1016/j.jmp.2017.09.005

Guitart-Masip, M., Bunzeck, N., Stephan, K. E., Dolan, R. J., & Düzel, E. (2010). Contextual novelty changes reward representations in the striatum. *Journal of Neuroscience*, *30*(5), 1721–1726. https://doi.org/10.1523/JNEUROSCI.5331-09.2010

Haaf, J. M., Rhodes, S., Naveh-Benjamin, M., Sun, T., Snyder, H. K., & Rouder, J. N. (2021). Revisiting the remember–know task: Replications of Gardiner and Java (1990). *Memory & Cognition*, *49*(1), 46–66. https://doi.org/10.3758/s13421-020-01073-x

Hennings, A. C., Lewis-Peacock, J. A., & Dunsmoor, J. E. (2021). Emotional learning retroactively enhances item memory but distorts source attribution. *Learning & Memory* , *28*(6), 178–186. https://doi.org/10.1101/lm.053371.120

Henson, R. N., Cansino, S., Herron, J. E., Robb, W. G. K., & Rugg, M. D. (2003). A familiarity signal in human anterior medial temporal cortex? *Hippocampus*, *13*(2), 301–304. https://doi.org/10.1002/hipo.10117

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–1326. https://doi.org/10.1002/hipo.20857

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1075–1090. https://doi.org/10.1037/xlm0000503

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, *1*(4), 304–309. https://doi.org/10.1038/1124

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition*, *154*, 151–164. https://doi.org/10.1016/j.cognition.2016.05.013

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, *2*(1), 105–112. https://doi.org/10.3758/BF03214414

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Kafkas, A., & Montaldi, D. (2015). Striatal and midbrain connectivity with the hippocampus selectively boosts memory for contextual novelty. *Hippocampus*, *25*(11), 1262–1273. https://doi.org/10.1002/hipo.22434

Kafkas, A., & Montaldi, D. (2018). How do memory systems detect and respond to novelty? *Neuroscience Letters*, *680*(July 2017), 60–68. https://doi.org/10.1016/j.neulet.2018.01.053

Kamiński, J., Mamelak, A. N., Birch, K., Mosher, C. P., Tagliati, M., & Rutishauser, U. (2018). Novelty-Sensitive Dopaminergic Neurons in the Human Substantia Nigra Predict Success of Declarative Memory Formation. *Current Biology*, *28*(9), 1333-1343.e4. https://doi.org/10.1016/j.cub.2018.03.024

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773. https://doi.org/10.2307/2291091

Kelley, M., & Nairne, J. (2001). Von Restorff Revisited: Isolation, Generation, and

Memory for Order. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*, 54–66. https://doi.org/10.1037//0278-7393.27.1.54

Kim, K., Yi, D. J., Raye, C. L., & Johnson, M. K. (2012). Negative effects of item repetition on source memory. *Memory and Cognition*, *40*(6), 889–901. https://doi.org/10.3758/s13421-012-0196-2

Kinsbourne, M., & George, J. (1974). The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(1), 63–69. https://doi.org/10.1016/S0022-5371(74)80031-9

Kleider, H. M., Pezdek, K., Goldinger, S. D., & Kirk, A. (2008). Schema-driven source misattribution errors: remembering the expected from a witnessed event. *Applied Cognitive Psychology*, *22*(1), 1–20. https://doi.org/10.1002/acp.1361

Kleiner, M., Brainard, D. H., & Pelli, D. (2007). What's new in Psychtoolbox-3? *PERCEPTION*, *36*(S), 14.

Kluen, L. M., Nixon, P., Agorastos, A., Wiedemann, K., & Schwabe, L. (2017). Impact of Stress and Glucocorticoids on Schema-Based Learning. *Neuropsychopharmacology*, *42*(6), 1254–1261. https://doi.org/10.1038/npp.2016.256

Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 699.

Konopka, A. E., & Benjamin, A. S. (2009). Schematic knowledge changes what judgments of learning predict in a source memory task. *Memory & Cognition*, *37*(1), 42–51. https://doi.org/10.3758/MC.37.1.42

Kormi-Nouri, R., Nilsson, L. G., & Ohta, N. (2005). The novelty effect: Support for the novelty-encoding hypothesis. *Scandinavian Journal of Psychology*, *46*(2), 133–143. https://doi.org/10.1111/j.1467-9450.2005.00443.x

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(5), 1365–1376. https://doi.org/10.1037/a0028147

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated.

Trends Cogn Sci, *20*(7), 512–534. https://doi.org/10.1016/j.tics.2016.05.004

Kumaran, D., & Maguire, E. A. (2006). An Unexpected Sequence of Events: Mismatch Detection in the Human Hippocampus. *PLoS Biology*, *4*(12), e424. https://doi.org/10.1371/journal.pbio.0040424

Kumaran, D., & Maguire, E. A. (2007). Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus*, *17*(9), 735–748. https://doi.org/10.1002/hipo.20326

Kumaran, D., & Maguire, E. A. (2009). Novelty signals: a window into hippocampal information processing. *Trends in Cognitive Sciences*, *13*(2), 47–54. https://doi.org/10.1016/j.tics.2008.11.004

Lampinen, J. M., Copeland, S., & Neuschatz, J. (2001). Recollections of Things Schematic: Room Schemas Revisited. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*, 1211–1222. https://doi.org/10.1037//0278-7393.27.5.1211

Lampinen, J. M., Faries, J. M., Neuschatz, J. S., & Toglia, M. P. (2000). Recollections of Things Schematic: The Influence of Scripts on Recollective Experience. *Applied Cognitive Psychology*, *14*(6), 543–554. https://doi.org/10.1002/1099-0720(200011/12)14:6<543::AID-ACP674>3.0.CO;2-K

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Lawrence, Z., & Peterson, D. (2016). Mentally walking through doorways causes forgetting: The location updating effect and imagination. *Memory*, *24*(1), 12–20. https://doi.org/10.1080/09658211.2014.980429

Lew, A. R., & Howe, M. L. (2017). Out of place, out of mind: Schema-driven false memory effects for object-location bindings. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(3), 404–421. https://doi.org/10.1037/xlm0000317

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

https://doi.org/10.1016/j.jesp.2013.03.013

Li, S., Cullen, W. K., Anwyl, R., & Rowan, M. J. (2003). Dopamine-dependent facilitation of LTP induction in hippocampal CA1 by exposure to spatial novelty. *Nature Neuroscience*, *6*(5), 526–531.

Lisman, J. E., & Grace, A. A. (2005). The Hippocampal-VTA Loop: Controlling the Entry of Information into Long-Term Memory. *Neuron*, *46*(5), 703–713. https://doi.org/10.1016/j.neuron.2005.05.002

Lisman, J. E., Grace, A. A., & Düzel, E. (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends in Neurosciences*, *34*(10), 536–547. https://doi.org/10.1016/j.tins.2011.07.006

Logie, M. R., & Donaldson, D. I. (2021). Do doorways really matter: Investigating memory benefits of event segmentation in a virtual learning environment. *Cognition*, *209*, 104578. https://doi.org/10.1016/J.COGNITION.2020.104578

Lopes da Cunha, P., Butavand, D., Chisari, L., Ballarini, F., & Viola, H. (2018). Exams at classroom have bidirectional effects on the long-term memory of an unrelated graphical task. *Npj Science of Learning*, *3*. https://doi.org/10.1038/s41539-018-0036-7

Lopes da Cunha, P., Villar, M. E., Ballarini, F., Tintorelli, R., & Viola, H. (2019). Spatial object recognition memory formation under acute stress. *Hippocampus*, *29*(6), 491–499. https://doi.org/10.1002/hipo.23037

Lovallo, W. R., Robinson, J. L., Glahn, D. C., & Fox, P. T. (2010). Acute effects of hydrocortisone on the human brain: An fMRI study. *Psychoneuroendocrinology*, *35*(1), 15–20. https://doi.org/10.1016/j.psyneuen.2009.09.010

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271. https://doi.org/10.1037/0033-295X.87.3.252

Mayes, A. R., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, *11*(3), 126–135. https://doi.org/https://doi.org/10.1016/j.tics.2006.12.003

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights From the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-

295X.102.3.419

McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It’s not surprising, it’s complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 422.

McFadyen, J., Nolan, C., Pinocy, E., Buteri, D., & Baumann, O. (2021). Doorways do not always cause forgetting: a multimodal investigation. *BMC Psychology*, *9*(1), 41. https://doi.org/10.1186/s40359-021-00536-3

McIntyre, C. K., McGaugh, J. L., & Williams, C. L. (2012). Interacting brain systems modulate memory consolidation. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1750–1762. https://doi.org/10.1016/j.neubiorev.2011.11.001

Mikell, C. B., Sheehy, J. P., Youngerman, B. E., McGovern, R. A., Wojtasiewicz, T. J., Chan, A. K., Pullman, S. L., Yu, Q., Goodman, R. R., Schevon, C. A., & McKhann, G. M. (2014). Features and timing of the response of single neurons to novelty in the substantia nigra. *Brain Research*, *1542*, 79–84. https://doi.org/10.1016/j.brainres.2013.10.033

Moncada, D., Ballarini, F., Martinez, M. C., & Viola, H. (2015). *The Behavioral Tagging Hypothesis and Its Implications for Long-Term Memory Formation BT - Synaptic Tagging and Capture: From Synapses to Behavior* (S. Sajikumar (ed.); pp. 231–259). Springer New York. https://doi.org/10.1007/978-1-4939-1761-7_14

Moncada, D., & Viola, H. (2007). Induction of Long-Term Memory by Exposure to Novelty Requires Protein Synthesis: Evidence for a Behavioral Tagging. *Journal of Neuroscience*, *27*(28), 7476–7481. https://doi.org/10.1523/JNEUROSCI.1083-07.2007

Montaldi, D., & Mayes, A. R. (2010). The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus*, *20*(11), 1291–1314. https://doi.org/10.1002/hipo.20853

Morey, R. D., Romeijn, J. W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, *72*, 6–18. https://doi.org/10.1016/j.jmp.2015.11.001

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs*. https://cran.r-project.org/package=BayesFactor

Morris, R. G. M. (2006). Elements of a neurobiological theory of hippocampal function:

the role of synaptic plasticity, synaptic tagging and schemas. *European Journal of Neuroscience*, *23*(11), 2829–2846. https://doi.org/10.1111/j.1460-9568.2006.04888.x

Moscovitch, M. (1995). Recovered consciousness: A hypothesis concerning modularity and episodic memory. *Journal of Clinical and Experimental Neuropsychology*, *17*(2), 276–290. https://doi.org/10.1080/01688639508405123

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic Memory and Beyond: The Hippocampus and Neocortex in Transformation. *Annual Review of Psychology*, *67*, 105–134. https://doi.org/10.1146/annurev-psych-113011-143733

Murty, V. P., Ballard, I. C., & Adcock, R. A. (2016). Hippocampus and Prefrontal Cortex Predict Distinct Timescales of Activation in the Human Ventral Tegmental Area. *Cerebral Cortex*, bhw005. https://doi.org/10.1093/cercor/bhw005

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227. https://doi.org/10.1016/S0959-4388(97)80010-4

Nielson, K. A., & Arentsen, T. J. (2012). Memory modulation in the classroom: Selective enhancement of college examination performance by arousal induced after lecture. *Neurobiology of Learning and Memory*, *98*(1), 12–16. https://doi.org/https://doi.org/10.1016/j.nlm.2012.04.002

Norman, K. A., Detre, G., & Polyn, S. M. (2008). Computational Models of Episodic Memory. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511816772.011

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford University Press.

Otten, L. J., Henson, R. N. A., & Rugg, M. D. (2001). Depth of processing effects on neural correlates of memory encoding: Relationship between findings from across- and within-task comparisons. *Brain*, *124*(2), 399–412. https://doi.org/10.1093/brain/124.2.399

Patel, P. D., Lopez, J. F., Lyons, D. M., Burke, S., Wallace, M., & Schatzberg, A. F. (2000). Glucocorticoid and mineralocorticoid receptor mRNA expression in squirrel monkey brain. *Journal of Psychiatric Research*, *34*(6), 383–392. https://doi.org/10.1016/S0022-3956(00)00035-2

Patil, A., Murty, V. P., Dunsmoor, J. E., Phelps, E. A., & Davachi, L. (2017). Reward retroactively enhances memory consolidation for related items. *Learning & Memory*, *24*(1), 65–69. https://doi.org/10.1101/lm.042978.116

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/https://doi.org/10.1163/156856897X00366

Pettijohn, K. A., & Radvansky, G. A. (2018a). Walking through doorways causes forgetting: active and passive interaction. *Journal of Cognitive Psychology*, *30*(8), 771–777. https://doi.org/10.1080/20445911.2018.1540492

Pettijohn, K. A., & Radvansky, G. A. (2018b). Walking through doorways causes forgetting: recall. *Memory*, *26*(10), 1430–1435. https://doi.org/10.1080/09658211.2018.1489555

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 587.

Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y., & Mendelsohn, A. (2018). Knowledge acquisition is governed by striatal prediction errors. *Nature Communications*, *9*(1), 1–14. https://doi.org/10.1038/s41467-018-03992-5

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

Poppenk, J., Köhler, S., & Moscovitch, M. (2010). Revisiting the novelty effect: When familiarity, not novelty, enhances memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(5), 1321–1330. https://doi.org/10.1037/a0019900

Prull, M. W. (2015). Adult age differences in memory for schema-consistent and schema-inconsistent objects in a real-world setting. *Aging, Neuropsychology, and Cognition*, *22*(6), 731–754. https://doi.org/10.1080/13825585.2015.1037821

Quent, J. A., Henson, R. N., & Greve, A. (2021). A predictive account of how novelty influences declarative memory. *Neurobiology of Learning and Memory*, *179*, 107382. https://doi.org/10.1016/j.nlm.2021.107382

Quent, J. A., McCullough, A. M., Sazma, M., Wolf, O. T., & Yonelinas, A. P. (2018).

Reward anticipation modulates the effect of stress-related increases in cortisol on episodic memory. *Neurobiology of Learning and Memory*, *147*, 65–73. https://doi.org/10.1016/J.NLM.2017.11.007

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition*, *34*(5), 1150–1156. https://doi.org/10.3758/BF03193261

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through Doorways Causes Forgetting: Further Explorations. *Quarterly Journal of Experimental Psychology*, *64*(8), 1632–1645. https://doi.org/10.1080/17470218.2011.571267

Radvansky, G. A., Tamplin, A. K., & Krawietz, S. A. (2010). Walking through doorways causes forgetting: Environmental integration. *Psychonomic Bulletin & Review*, *17*(6), 900–904. https://doi.org/10.3758/PBR.17.6.900

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*(1), 89–102.

Ramirez Butavand, D., Hirsch, I., Tomaiuolo, M., Moncada, D., Viola, H., & Ballarini, F. (2020). Novelty Improves the Formation and Persistence of Memory in a Naturalistic School Scenario. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00048

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, *4*, 193. https://doi.org/10.1038/nrn1052

Raz, N. (2005). Ageing and the Brain. In *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd. https://doi.org/10.1038/npg.els.0004063

Redondo, R. L., & Morris, R. G. M. (2011). Making memories last: The synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience*, *12*(1), 17–30. https://doi.org/10.1038/nrn2963

Reggev, N., Sharoni, R., & Maril, A. (2018). Distinctiveness Benefits Novelty (and Not Familiarity), but Only Up to a Limit: The Prior Knowledge Perspective. *Cognitive Science*, *42*(1), 103–128. https://doi.org/10.1111/cogs.12498

Reichardt, R., Polner, B., & Simor, P. (2020). Novelty Manipulations, Memory

Performance, and Predictive Coding: the Role of Unexpectedness. *Frontiers in Human Neuroscience*, *14*(April), 1–11. https://doi.org/10.3389/fnhum.2020.00152

Rescorla, R. A., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory: Vol. Vol. 2*.

Reul, J. M. H. M., & de Kloet, E. R. (1985). Two Receptor Systems for Corticosterone in Rat Brain: Microdistribution and Differential Occupation. *Endocrinology*, *117*(6), 2505–2511. https://doi.org/10.1210/endo-117-6-2505

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*(4), 613–643. https://doi.org/10.1080/15326900701399913

Richmond, L. L., & Zacks, J. M. (2017). Constructing Experience: Event Models from Perception to Action. *Trends in Cognitive Sciences*, *21*(12), 962–980. https://doi.org/10.1016/j.tics.2017.08.005

Richter-Levin, G., & Akirav, I. (2003). Emotional tagging of memory formation—in the search for neural mechanisms. *Brain Research Reviews*, *43*(3), 247–256. https://doi.org/10.1016/j.brainresrev.2003.08.005

Ritchey, M., McCullough, A. M., Ranganath, C., & Yonelinas, A. P. (2017). Stress as a mnemonic filter: Interactions between medial temporal lobe encoding processes and post-encoding stress. *Hippocampus*, *27*(1), 77–88. https://doi.org/10.1002/hipo.22674

Roig, M., Nordbrandt, S., Geertsen, S. S., & Nielsen, J. B. (2013). The effects of cardiovascular exercise on human memory: A review with meta-analysis. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1645–1666. https://doi.org/10.1016/j.neubiorev.2013.06.012

Roig, M., Thomas, R., Mang, C. S., Snow, N. J., Ostadan, F., Boyd, L. A., & Lundbye-Jensen, J. (2016). Time-dependent effects of cardiovascular exercise on memory. *Exercise and Sport Sciences Reviews*, *44*(2), 81–88.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2019). Reward prediction errors create event boundaries in memory. *BioRxiv*, 725440. https://doi.org/10.1101/725440

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition*, *203*, 104269. https://doi.org/https://doi.org/10.1016/j.cognition.2020.104269

Sánchez, M. M., Young, L. J., Plotsky, P. M., & Insel, T. R. (2000). Distribution of Corticosteroid Receptors in the Rhesus Brain: Relative Absence of Glucocorticoid Receptors in the Hippocampal Formation. *The Journal of Neuroscience, 20*(12), 4657–4668. https://doi.org/10.1523/JNEUROSCI.20-12-04657.2000

Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In *Memory Systems 1994* (pp. 1–38). MIT Press.

Schmidt, S. R., & Schmidt, C. R. (2017). Revisiting von Restorff's early isolation effect. *Memory and Cognition*, *45*(2), 194–207. https://doi.org/10.3758/s13421-016-0651-6

Schomaker, J. (2019). Unexplored territory: Beneficial effects of novelty on memory. *Neurobiology of Learning and Memory*. https://doi.org/https://doi.org/10.1016/j.nlm.2019.03.005

Schomaker, J., van Bronkhorst, M. L. V, & Meeter, M. (2014). Exploring a novel environment improves motivation and promotes recall of words. *Frontiers in Psychology*, *5*(AUG), 1–6. https://doi.org/10.3389/fpsyg.2014.00918

Schomaker, J., & Wittmann, B. C. (2021). Effects of active exploration on novelty-related declarative memory enhancement. *Neurobiology of Learning and Memory*, *179*, 107403. https://doi.org/https://doi.org/10.1016/j.nlm.2021.107403

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schott, B. H., Sellner, D. B., Lauer, C. J., Habib, R., Frey, J. U., Guderian, S., Heinze, H. J., & Düzel, E. (2004). Activation of midbrain structures by associative novelty and the formation of explicit memory in humans. *Learning and Memory*, *11*(4), 383–

387. https://doi.org/10.1101/lm.75004

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Schultz, W., & Dickinson, A. (2000). Neuronal Coding of Prediction Errors. *Annual Review of Neuroscience*, *23*(1), 473–500. https://doi.org/10.1146/annurev.neuro.23.1.473

Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: An update and integration. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1740–1749. https://doi.org/10.1016/j.neubiorev.2011.07.002

Scoville, W. B., & Milner, B. (1957). LOSS OF RECENT MEMORY AFTER BILATERAL HIPPOCAMPAL LESIONS. *Journal of Neurology, Neurosurgery & Psychiatry*, *20*(1), 11–21. https://doi.org/10.1136/jnnp.20.1.11

Seel, S. V., Easton, A., McGregor, A., Buckley, M. G., & Eacott, M. J. (2019). Walking through doorways differentially affects recall and familiarity. *British Journal of Psychology*, *110*(1), 173–184. https://doi.org/10.1111/bjop.12343

Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters*, *680*, 39–53. https://doi.org/10.1016/j.neulet.2018.05.006

Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of acute stress on episodic memory: A meta-analysis and integrative review. *Psychological Bulletin*, *143*(6), 636–675. https://doi.org/10.1037/bul0000100

Shin, Y. S., Masís-Obando, R., Keshavarzian, N., Dáve, R., & Norman, K. A. (2021). Context-dependent memory effects in two immersive virtual reality environments: On Mars and underwater. *Psychonomic Bulletin & Review*, *28*(2), 574–582. https://doi.org/10.3758/s13423-020-01835-3

Simonsohn, U. (2018). Two Lines: A Valid Alternative to the Invalid Testing of U-Shaped Relationships With Quadratic Regressions. *Advances in Methods and Practices in Psychological Science*, *1*(4), 538–555. https://doi.org/10.1177/2515245918805755

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, *45*(2), 560–575.

https://doi.org/10.3758/s13428-012-0259-0

Song, Z., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2011). Impaired capacity for familiarity after hippocampal damage. *Proceedings of the National Academy of Sciences*, *108*(23), 9655 LP – 9660. https://doi.org/10.1073/pnas.1107247108

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195–231. https://doi.org/10.1037/0033-295X.99.2.195

Straube, T., Korz, V., Balschun, D., & Frey, J. U. (2003). Requirement of β-adrenergic receptor activation and protein synthesis for LTP-reinforcement by novelty in rat dentate gyrus. *Journal of Physiology*, *552*(3), 953–960. https://doi.org/10.1113/jphysiol.2003.049452

Straube, T., Korz, V., & Frey, J. U. (2003). Bidirectional modulation of long-term potentiation by novelty-exploration in rat dentate gyrus. *Neuroscience Letters*, *344*(1), 5–8. https://doi.org/10.1016/S0304-3940(03)00349-5

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236–257. https://doi.org/10.1037/a0015631

Tse, D., Langston, R. R. F., Kakeyama, M., Bethus, I., Spooner, P. a, Wood, E. R., Witter, M. P., & Morris, R. G. M. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82. https://doi.org/10.1126/science.1135935

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., & Morris, R. G. M. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science*, *333*(6044), 891–895.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Tulving, E. (1995). Organization of memory: quo vadis? In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 839–853). The MIT Press.

Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review*, *2*(3), 387–390. https://doi.org/10.3758/BF03210977

Tulving, E., Markowitsch, H. J., Craik, F. I. M., Habib, R., & Houle, S. (1996). Novelty and Familiarity Activations in PET Studies of Memory Encoding and Retrieval. *Cerebral Cortex*, *6*(1), 71–79. https://doi.org/10.1093/cercor/6.1.71

van Buuren, M., Kroes, M. C. W., Wagner, I. C., Genzel, L., Morris, R. G. M., & Fernández, G. (2014). Initial investigation of the effects of an experimentally learned schema on spatial associative memory in humans. *Journal of Neuroscience*, *34*(50), 16662–16670.

van der Linden, M., Berkers, R. M. W. J., Morris, R. G. M., & Fernández, G. (2017). Angular gyrus involvement at encoding and retrieval is associated with durable but less specific memories. *The Journal of Neuroscience*, *37*(39), 3603–3616. https://doi.org/10.1523/JNEUROSCI.3603-16.2017

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Vogel, S., Kluen, L. M., Fernández, G., & Schwabe, L. (2018). Stress leads to aberrant hippocampal involvement when processing schema-related information. *Learning & Memory*, *25*(1), 21–30. https://doi.org/10.1101/lm.046003.117

Vogel, S., & Schwabe, L. (2016). Learning and memory under stress: implications for the classroom. *Npj Science of Learning*, *1*, 16011. https://doi.org/10.1038/npjscilearn.2016.11

von Restorff, H. (1933). Analyse yon Vorgängen im Spurenfeld. *Psychological Research*, *61*(1912), 299–342.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/https://doi.org/10.1016/j.cogpsych.2009.12.001

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. J. (2011). *Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011).* http://psycnet.apa.org/journals/psp/100/3/426/

Wang, S.-H., Redondo, R. L., & Morris, R. G. M. (2010). Relevance of synaptic tagging and capture to the persistence of long-term potentiation and everyday spatial memory. *Proceedings of the National Academy of Sciences*, *107*(45), 19537–19542. https://doi.org/10.1073/pnas.1008638107

Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, *20*(1), 6–11.

Winocur, G., & Moscovitch, M. (2011). Memory Transformation and Systems Consolidation. *Journal of the International Neuropsychological Society*, *17*(05), 766–780. https://doi.org/10.1017/S1355617711000683

Wittmann, B. C., Bunzeck, N., Dolan, R. J., & Düzel, E. (2007). Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *NeuroImage*, *38*(1), 194–202. https://doi.org/10.1016/j.neuroimage.2007.06.038

Wixted, J. T. (2004). The Psychology and Neuroscience of Forgetting. *Annual Review of Psychology*, *55*(1), 235–269. https://doi.org/10.1146/annurev.psych.55.090902.141555

Wolf, O. T. (2008). The influence of stress hormones on emotional memory: Relevance for psychopathology. *Acta Psychologica*, *127*(3), 513–531. https://doi.org/10.1016/j.actpsy.2007.08.002

Wolf, O. T. (2009). Stress and memory in humans: Twelve years of progress? *Brain Research*, *1293*, 142–154. https://doi.org/10.1016/j.brainres.2009.04.013

Worthen, J., & Hunt, R. R. (2006). *Distinctiveness and memory*. Oxford University Press.

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nature Reviews Neuroscience*, *20*(6), 364–375. https://doi.org/10.1038/s41583-019-0150-4

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

7 APPENDICES

Appendix for chapter 2

Table 7.1: List of priors for models reported in this chapter.

	Experiment	**Exp. 1**			**Exp. 2**			**Exp. 3a**			**Exp. 3b**		
Measure	Predictor	df	μ	σ	df	μ	σ	df	μ	σ	df	μ	σ
Recall	Intercept	7	0	10	20.11	-0.95	0.42	68.3	-1.25	0.35	61.27	-0.67	0.26
	Linear	7	0	1	68.18	0.1	0.39	114.77	0.41	0.25	104.22	0.19	0.17
	Quadratic	7	0	1	63.49	1.88	0.84	118.29	1.54	0.55	132.09	1.15	0.4
3AFC	Intercept	7	0	10	16.92	0.91	0.37	46.04	0.68	0.27	52.23	0.72	0.2
	Linear	7	0	1	72.54	0.29	0.4	114.87	0.75	0.25	116.12	0.59	0.18
	Quadratic	7	0	1	36.25	0.63	0.69	69.6	1.17	0.5	101.84	1.14	0.4
Recollection (independence/ redundancy)	Intercept				7	0	10	24.06	-0.48	0.29	63.4	-0.5	0.21
	Linear				7	0	1	104.49	0.29	0.26	126.96	0.03	0.17
	Quadratic				7	0	1	104.51	1.48	0.57	142.09	1.47	0.4
Recollection (exclusivity)	Intercept				7	0	10	23.45	0.37	0.37	53.29	0.35	0.25
	Linear				7	0	1	115.15	0.32	0.32	112.25	0.03	0.21
	Quadratic				7	0	1	55.59	1.32	0.69	113.73	1.44	0.49
Familiarity (independence, exclusivity)	Intercept				7	0	10	23.32	-0.01	0.23	39.65	-0.01	0.19
	Linear				7	0	1	87.19	0.1	0.3	82.92	-0.04	0.22
	Quadratic				7	0	1	49.03	0.1	0.61	67.99	0.11	0.48
Familiarity (redundancy)	Intercept				7	0	10	26.43	0.96	0.26	43.73	0.93	0.19
	Linear				7	0	1	111.07	0.19	0.28	84.5	-0.05	0.19
	Quadratic				7	0	1	76.77	0.96	0.58	116.98	1.01	0.43

Table 7.2: List of object locations

Object #	Name	Locations Set 1	Set 2	Set 3 (111)	Set 4 (246)	Set 5 (388)	Set 6 (498)	Set 7 (848)
1	microwave	14	19	13	13	11	10	19
2	kitchen roll	8	17	11	16	4	19	2
3	saucepan	3	2	15	18	15	12	13
4	toaster	12	4	20	1	8	1	9
5	bowl of fruits	13	11	6	16	18	16	1
6	tea pot	18	5	5	7	2	15	4
7	knife	20	13	10	14	6	13	14
8	mixer	9	18	12	3	14	14	17
9	bread	4	15	19	11	16	4	16
10	glass jug	15	7	8	4	7	8	6
11	mug	17	3	3	19	20	3	11
12	dishes	6	10	9	12	10	20	10
13	towels	7	14	17	10	5	18	5
14	toy	11	12	2	6	1	9	15
15	books	2	1	4	8	19	7	7
16	umbrella	1	8	18	17	12	5	8
17	hat	16	6	1	2	17	11	20
18	helmet	5	9	16	20	3	6	12
19	calendar	10	20	14	5	13	17	3
20	fan	19	16	7	9	9	2	18

7.1 Original analysis of remember/know data

Remember/familiar judgements were initially analysed in line with pre-registered analysis of the mean expectancy rating for remember and familiar judgments, but further simulation showed that this trial-averaged analysis is biased by boundary effects on expectancy values (see https://jaquent.github.io/post/a-u-shape-that-appears-as-a-linear-correlation-when-averaged/). The correlation between expectancy and the average number of remember response per object was, r = -0.41, t(18) = -1.9204, p = .070, while the correlation between expectancy and the average number of familiar response was, r = 0.40, t(18) = 1.8752, p = .08.

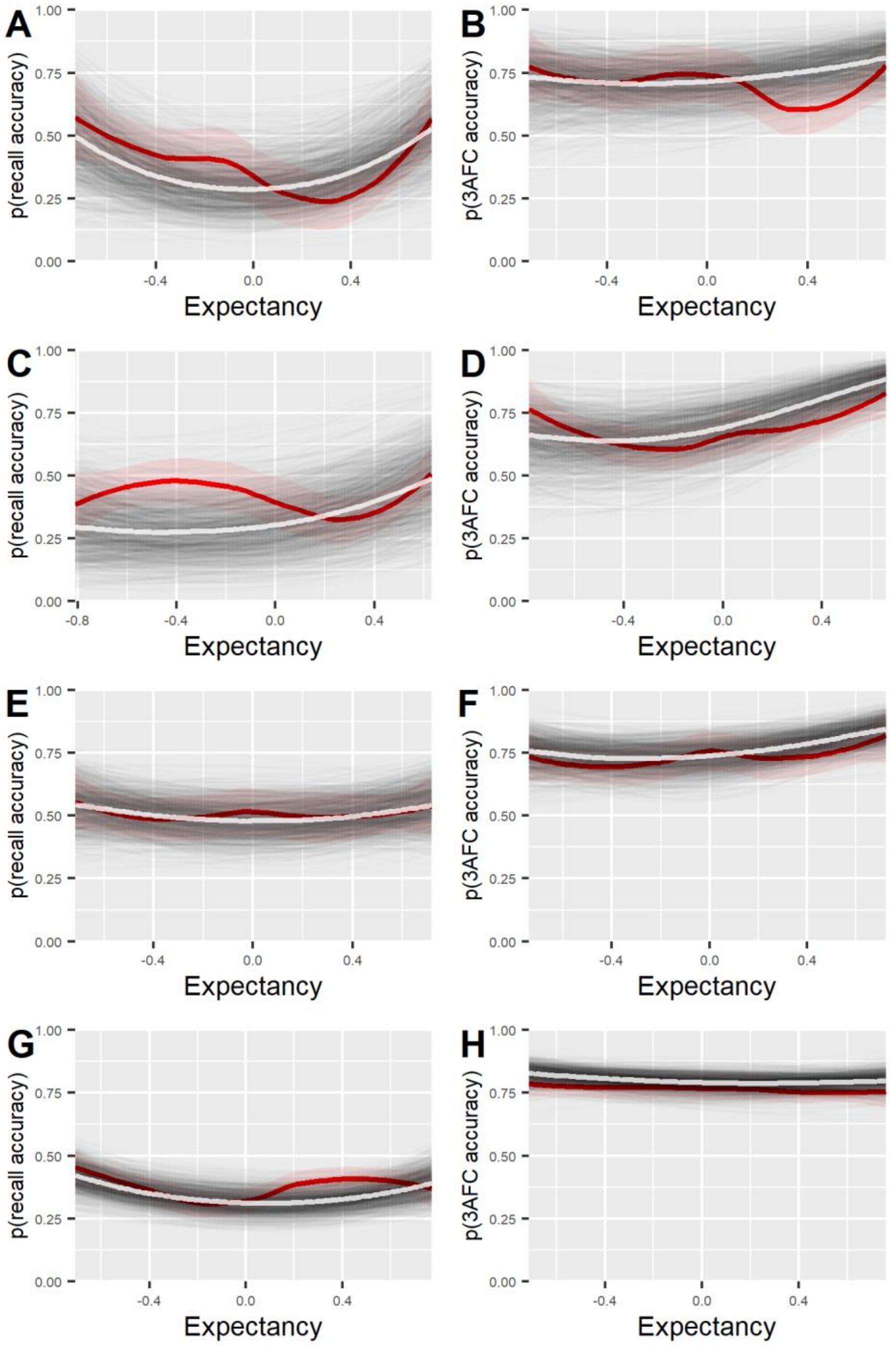

Figure 7.1: Accuracy results (non-propagated): Recall (left panels) and recognition (right panels) performance plotted across expectancy ratings for all four Experiments 1, 2, 3a and 3b (rows). The red line shows locally weighted smoothing (loess) of data to illustrate how well the fitted 2nd-order polynomial model using non-propagated evidence (white line) represents the data. The thin grey lines show 1,000 randomly selected fits from the posterior distributions, illustrating the uncertainty of the fit. Expectancy ratings originally ranged from -100 to +100, but were scaled to have SD = 0.5 in order to enable standard priors.

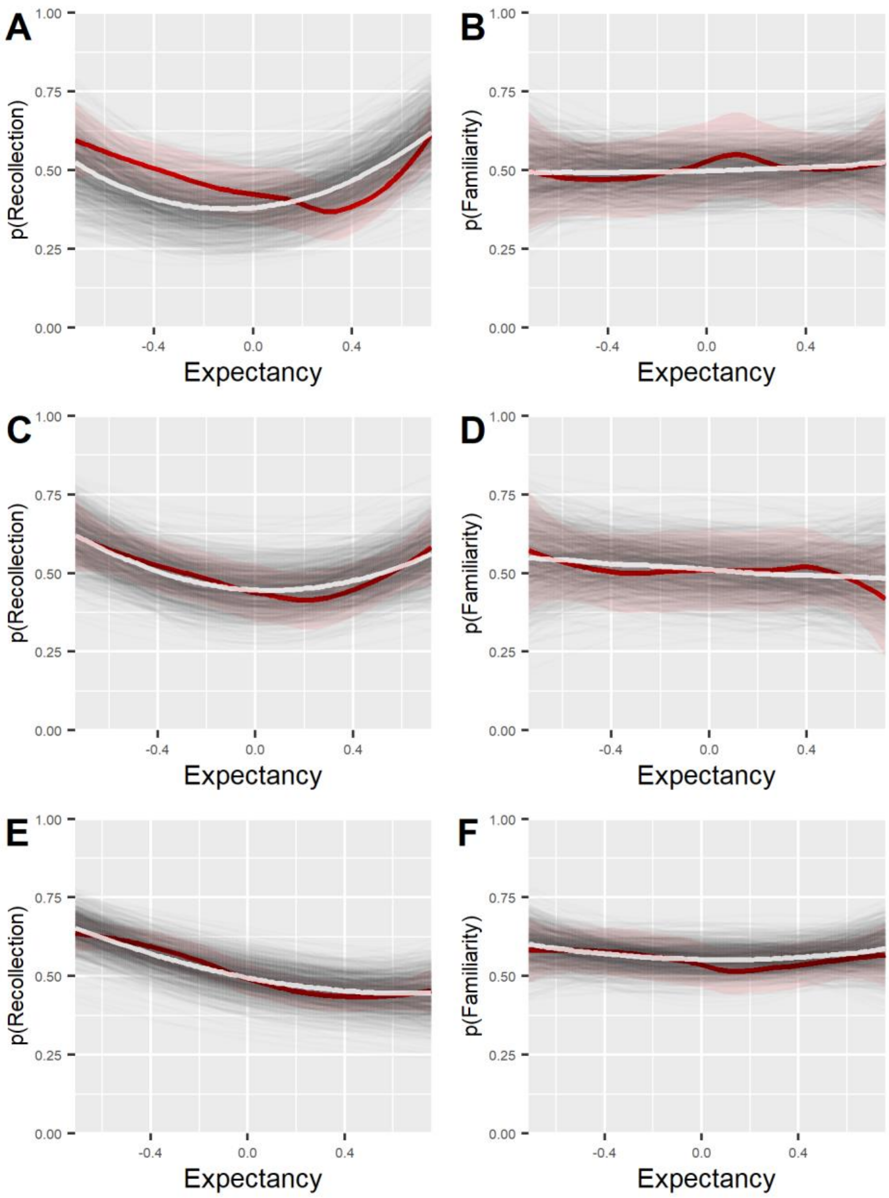

Figure 7.2: Recollection & familiarity results (non-propagated): 3AFC recognition performance based on recollection (under independent or redundant scoring, left panel) and familiarity (under independent or exclusive scoring, right panel) estimates across Experiment 2, Experiment 3a and Experiment 3b (rows). The red line shows locally weighted smoothing (loess) of data to illustrate how well the average model (the white line) represents the data. The thin grey lines show 1,000 randomly selected fits from the posterior distributions, to illustrate the uncertainty of the average model. Expectancy originally ranged from -100 to +100, but were scaled to have a SD = 0.5 in order to enable standard priors.

Table 7.3: Interrupted regression results for recall: A U-shape was found for breaking point 6 for recall as the CI and the BFs (order restricted) indicated that both slopes are opposites signs and different from zero.

Breaking point #	Breaking point position	β_1	2.5%	97.5%	BF_{10}	BF_{10} (OR)	β_2	2.5%	97.5%	BF_{10}	BF_{10} (OR)
1	-0.73	-0.01	-2.36	2.29	0.99	0.98	0.07	-0.15	0.29	0.13	0.19
2	-0.58	0.27	-1.63	2.3	0.93	0.74	0.26	-0.01	0.54	0.8	1.54
3	-0.42	-0.01	-1.3	1.28	0.63	0.64	0.52	0.19	0.86	17.1	34.16
4	-0.26	-0.85	-1.8	0.08	2.41	4.64	0.53	0.15	0.92	7.06	14.06
5	-0.1	-0.61	-1.23	0.01	2.08	4.05	0.72	0.23	1.22	13.59	27.11
6	0.06	-0.78	-1.3	-0.28	27.46	54.89	0.69	0.06	1.34	3.18	6.27
7	0.22	-0.38	-0.78	0.01	1.16	2.25	1.35	0.46	2.27	35.93	71.75
8	0.38	-0.35	-0.67	-0.03	1.71	3.37	1.78	0.36	3.27	16.2	32.2
9	0.54	-0.31	-0.58	-0.04	1.67	3.3	1.15	-0.72	3.45	1.65	2.89
10	0.69	-0.19	-0.41	0.03	0.44	0.84	0.02	-2.31	2.35	1.05	1.05

Table 7.4: Interrupted regression results for 3AFC: No U-shape was detected for 3AFC as the CI and the BFs (order restricted) indicated that there is enough evidence that both slopes are opposites signs and different from zero.

Breaking point #	Breaking point position	β_1	2.5%	97.5%	BF_{10}	BF_{10} (OR)	β_2	2.5%	97.5%	BF_{10}	BF_{10} (OR)
1	-0.73	0.01	-2.31	2.3	1.05	1.04	0.27	0.03	0.51	1.42	2.81
2	-0.58	-0.63	-2.92	1.34	1.05	1.52	0.4	0.12	0.68	5.88	11.73
3	-0.42	-0.54	-1.91	0.8	0.86	1.34	0.55	0.21	0.89	29.23	58.41
4	-0.26	-0.84	-1.77	0.12	2.06	3.93	0.5	0.1	0.92	4.28	8.48
5	-0.1	-0.51	-1.16	0.14	1.08	2.02	0.55	0.03	1.08	2.38	4.66
6	0.06	-0.31	-0.82	0.21	0.49	0.87	0.75	0.11	1.42	4.36	8.63
7	0.22	-0.01	-0.42	0.41	0.2	0.21	1.64	0.69	2.61	116.52	232.88
8	0.38	-0.13	-0.46	0.21	0.22	0.33	2.24	0.71	3.95	44.31	88.42
9	0.53	-0.19	-0.47	0.09	0.31	0.56	1.1	-0.78	3.57	1.5	2.57
10	0.69	-0.09	-0.33	0.14	0.16	0.24	0.06	-2.28	2.5	0.99	1.04

Table 7.5: Interrupted regression for recollection: A U-shape was found for breaking points 6 - 9 in recollection as the CI and the BFs (order restricted) indicated that both slopes are opposites signs and different from zero.

Breaking point #	Breaking point position	β_1	2.5%	97.5%	BF_{10}	BF_{10} (OR)	β_2	2.5%	97.5%	BF_{10}	BF_{10}(OR)
1	-0.73	0.01	-2.3	2.38	1	0.99	-0.35	-0.54	-0.15	20.37	0.01
2	-0.57	0.01	-1.8	1.88	0.87	0.86	-0.15	-0.42	0.11	0.24	0.06
3	-0.42	-0.09	-1.41	1.19	0.64	0.71	0.08	-0.23	0.39	0.17	0.24
4	-0.26	-0.77	-1.7	0.17	1.75	3.33	0.13	-0.24	0.49	0.23	0.35
5	-0.1	-0.72	-1.35	-0.11	4.25	8.4	0.48	0	0.95	1.67	3.26
6	0.06	-0.85	-1.36	-0.35	46.02	91.97	0.84	0.21	1.48	12.25	24.42
7	0.22	-0.84	-1.23	-0.45	309.78	619.55	1.22	0.37	2.1	20.27	40.45
8	0.38	-0.82	-1.15	-0.5	4342.91	8685.82	1.67	0.33	3.15	12.91	25.63
9	0.53	-0.72	-0.99	-0.45	3570.19	7140.37	1.98	-0.13	4.77	4.39	8.45
10	0.69	-0.62	-0.84	-0.39	3407.16	6814.31	0.14	-2.22	2.6	0.97	1.07

Table 7.6: Interrupted regression results for familiarity: No U-shape was detected for familiarity scored following independence assumptions as the CI and the BFs (order restricted) indicated that both slopes are not opposites signs and different from zero.

Breaking point #	Breaking point position	β_1	2.5%	97.5%	BF_{10}	BF_{10} (OR)	β_2	2.5%	97.5%	BF_{10}	BF_{10} (OR)
1	-0.73	0.01	-2.45	2.43	1	0.98	-0.18	-0.42	0.08	0.31	0.05
2	-0.57	0.07	-2.04	2.09	0.92	0.86	-0.05	-0.39	0.28	0.17	0.13
3	-0.42	0.21	-1.28	1.75	0.76	0.59	0.13	-0.26	0.5	0.23	0.34
4	-0.26	-0.34	-1.45	0.74	0.61	0.9	0.24	-0.22	0.72	0.38	0.64
5	-0.1	-0.76	-1.56	0.03	2.38	4.62	0.2	-0.36	0.79	0.35	0.53
6	0.06	-0.56	-1.16	0.03	1.72	3.34	0.46	-0.27	1.21	0.76	1.35
7	0.22	-0.45	-0.93	0.02	1.37	2.65	0.94	-0.06	1.98	2.86	5.53
8	0.38	-0.61	-1.01	-0.22	16.76	33.48	0	-1.4	1.39	0.68	0.67
9	0.53	-0.37	-0.69	-0.05	1.95	3.84	0.08	-1.77	2.04	0.89	0.94
10	0.69	-0.27	-0.54	0.02	0.82	1.6	0.08	-2.3	2.52	1.04	1.09